

Detection of sparse additive functions

Ghislaine Gayraud*

*Université de Technologie de Compiègne & CREST,
BP 20529, 60205 Compiègne, France
email: ghislaine.gayraud@utc.fr*

and

Yuri Ingster†

*St. Petersburg State Electrotechnical University, 5,
Prof. Popov str., 197376 St. Petersburg, Russia
email: yurii_ingster@mail.ru*

Abstract: We study the problem of detection of high-dimensional signal functions in the Gaussian white noise model. We assume that, in addition to a smoothness assumption, the signal function has an additive sparse structure. The detection problem is expressed in terms of a nonparametric hypothesis testing problem and is solved using asymptotically minimax approach. We provide minimax test procedures that are adaptive in the sparsity parameter in the high sparsity case. We extend some known results related to the detection of sparse high-dimensional vectors to the functional case. In particular, our derivation of asymptotic detection rates is based on same detection boundaries as in the vector case.

AMS 2000 subject classifications: 62H15, 60G15, 62G10, 62G20, 60C20.

Keywords and phrases: High-dimensional setting, sparsity, asymptotic minimax approach, detection boundary, Gaussian white noise model.

Contents

1	Introduction	1
2	Detection boundaries in a vectorial Gaussian model	4
3	Transformation of the statistical testing problem	4
4	Extremal problem	6
5	Main results	9
5.1	Moderate sparsity	10
5.2	High sparsity	11
6	Extended problem	12
7	Proofs	13
7.1	Properties of test statistics	14
7.2	Upper bound	17
7.3	Lower Bound	21
7.4	Appendix	28
7.4.1	Proof of Lemma 7.2.	28
7.4.2	Proof of Lemma 7.3.	29
	References	31

1. Introduction

Over the past years, boosted by applications and computer performance, problems in high-dimensions have been explored in a number of statistical studies. If no additional structure is assumed, high-dimensional data processing suffers from some intrinsic difficulties such as the curse of dimensionality that results in a loss in the efficiency of statistical procedures, and inconsistency

*Ghislaine Gayraud's research was partially supported by the ANR-blanc 'SP Bayes'

†Yuri Ingster's research was partially supported by RFBR Grant 11-01-00577 and by Grant NSh-4472.2010.1

of classical statistical procedures – even in the linear regression model – unless the dimension of variables is less than the sample size.

In order to overcome the curse of dimensionality in a nonparametric framework, where typical functional classes are Sobolev, Holder, or Besov balls, some additional conditions, including additivity or tensor product structure, are assumed, see, for instance, [20, 6, 18, 14, 15, 16] and references therein. Even if one of these conditions is assumed, yet it is required that the sample size is to be larger than the data dimension. One way to free oneself from the latter condition is to impose an additional sparsity constraint.

In this paper we focus on the problem of detection of high-dimensional signal functions in the Gaussian white noise model. To avoid difficulties stemming from high-dimensional settings, we suppose that a signal function satisfies an additional structural condition. Specifically, it is assumed to be sparse additive. This means that a high-dimensional function of interest is a sum of few univariate functions. Formally, we consider an d -dimensional ($d \in \mathbb{N}$ and $d > 0$) Gaussian white noise model

$$dX(t) = f(t)dt + \epsilon dW(t), \quad t \in [0, 1]^d, \quad (1.1)$$

where $W(t)$ is the Wiener process, $\epsilon > 0$ is the noise level, and f , the quantity of interest, is the signal function. The additive sparse structure means that f is the sum of d univariate functions f_j :

$$f(t) = \sum_{j=1}^d \xi_j f_j(t_j), \quad t_j \in [0, 1], \quad (1.2)$$

where the ξ_j 's are unknown but deterministic taking their values in $\{0, 1\}$: “0” means that the j th component f_j is non active whereas “1” means that f_j is active. Denote by K the positive number of active components, that is, $K = \sum_{j=1}^d \xi_j$, and assume that $K = d^{1-b}$, where $b \in (0, 1)$ is

the *sparsity index*. If d^{1-b} is not an integer then take K as its integer part. Denote by $\mathcal{F}_{d,b}$ the functional class of additive sparse signals f of the form (1.2) with $K = d^{1-b}$ active components and d^b non-active components. Model (1.1) with the sparse additive structure (1.2) is a natural generalization of the sparse linear model: the nonparametric nature of the problem suggests to consider more flexible models.

There is a huge statistical literature on estimation in sparse models, see, for instance, [1, 2, 3] and references therein. In particular, there are many works related to the well-known Lasso introduced by Tibshirani [21] in 1996. There are also a number of papers that deal with nonparametric estimation in sparse additive models. For a complete review of these topics, we refer to [19], where minimax estimation rates in sparse additive models are obtained, to [5], where the Lasso-type estimate in sparse additive models is studied, and to [20], where various structural assumptions on models in high dimensions are discussed.

Back to our study, the detection problem at hand can be expressed in terms of a nonparametric hypothesis testing problem with the null hypothesis stating that “the signal is a constant”, and “there is no signal” being a particular case of the null hypothesis. In order to specify an alternative hypothesis, recall that, within the minimax framework, it is impossible to detect signal functions that are “too close” to the null one, as well as to test the null and alternative hypotheses for too large alternative classes. Therefore, we are interested in the following nonparametric hypothesis testing problem:

$$H_0 : f = \text{const}_0 \quad \text{versus} \quad H_1 : f = \text{const}_1 + f^1, \quad f^1 \in \mathcal{F}_d(\tau, r_\epsilon, b), \quad (1.3)$$

where

$$\begin{cases} \text{const}_0, \text{const}_1 \text{ are some constants,} \\ \mathcal{F}_d(\tau, r_\epsilon, b) = \{f \in \mathcal{F}_{d,b} : \forall j, f_j \in \tilde{S}_\tau \text{ and } \|f_j\|_2 \geq r_\epsilon\}, \quad \tau > 0, r_\epsilon > 0, \\ \tilde{S}_\tau = \{f \in L_2([0, 1]) : \int_0^1 f(t)dt = 0, \|f\|_2^{(\tau)} \leq 1\}. \end{cases}$$

The L_2 -norm $\|\cdot\|_2$ is used to separate the nonparametric alternative from the null hypothesis. The functional class \tilde{S}_τ is the Sobolev ball, expressed via the Sobolev semi-norm $\|\cdot\|_2^{(\tau)}$, that contains τ -smooth functions, which are assumed 1-periodic and orthogonal to a constant. Due to the periodic constraints, it is possible to express $\|\cdot\|_2^{(\tau)}$ in terms of Fourier coefficients; this will be done in Section 2. The quantity τ is the smoothness parameter. Both the smoothness condition and the separation condition between H_0 and H_1 are expressed in terms of the components f_j that are linked to the whole signal f via (1.2): each active component f_j is smooth and is separated from the null hypothesis in the L_2 -norm by a positive value r_ϵ .

In Section 6, we generalize the hypothesis testing problem (1.3) by considering a more general class of alternatives that consists of signals f equal, up to a constant, to a function $f^1 \in \mathcal{F}_{d,b}$, which is separated from the null hypothesis in the $L_2([0, 1]^K)$ -norm, and whose smoothness is expressed in terms of the whole function f .

For these two hypothesis testing problems, the main questions are: *what are the separation rates in the problem, i.e., what are the asymptotics for the minimal r_ϵ such that one can distinguish between H_0 and H_1 ? And, also, what are the optimal test procedures that provide distinguishability?*

To answer these questions, we use asymptotically minimax approach that provides detection boundaries or distinguishability conditions, i.e., necessary and sufficient conditions for the possibility of successful detection; these detection boundaries yield asymptotics for the minimal r_ϵ separating the areas of distinguishability and non-distinguishability (between H_0 and H_1). The asymptotics for the minimal values of r_ϵ are called either the (minimax) separation rates or the minimax rates of testing; in the present paper, the separation rates are denoted by r_ϵ^* .

In connection with the current study, a number of works on detection and classification boundaries in Gaussian sequence models could be mentioned, see, for example, [7, 8, 9, 13, 12, 4, 15, 16, 11]. Also, in [17], rather than considering a Gaussian sequence model, the authors generalize the problem of finding a detection boundary in the linear regression model. Another paper [10] deals with the signal detection problem in a multichannel model in the functional framework. At the end of the next paragraph, we explain what are the differences between the results in [10] and our study.

The main contribution of this paper consists of extending the results on detection boundaries obtained for d -dimensional sparse Gaussian vectors, see, for instance, [12], to the functional case. In particular, we obtain the same detection boundaries as in the vectorial case. However, in the case of high sparsity when $b > 1/2$, an additional assumption on the growth of d as a function of ϵ is required. Distinguishability is possible when the sum of the type I error probability and the maximum over alternatives of the type II error probability vanishes asymptotically, and distinguishability is not possible when this sum tends to one. Boundary conditions depend on the quantity $a(r_\epsilon) = a(r_\epsilon, d, \tau)$, which is a solution of a certain extremal problem stated in Section 4. In the vectorial case, the quantity $a(r_\epsilon)$ corresponds to the energy of a signal (see [12] and [10]). In the functional case, this quantity characterizes the distinguishability in a one-variable hypotheses testing problem. The minimax separation rates obtained in this paper depend on the value of b : for large b they are worse than for small b . Such a behaviour is expected because, with large b , only few components are active, and hence the problem of distinguishing between the alternative and null hypothesis becomes more difficult.

For the most difficult case of $b \in (1/2, 1)$, not only separation rates, but also sharp separation rates, that include both rates and constants, are obtained. We also provide optimal test procedures for which minimax rates of testing are achieved asymptotically. Depending on the value of b , we propose two types of test procedures: one is of a χ^2 type, the other one is related to a Higher-Criticism statistic introduced in [4] and based on the Tukey's ideas. In the case of $b \in (1/2, 1)$, our test procedure is adaptive in the sparsity index b , see Remark 5.3.

In the paper [10], which is focused on a similar problem of multichannel signal detection, the optimal rates are obtained. In our study, we obtain sharp separation rates for $b \in (1/2, 1)$. The main difference between the study of [10] and our work is in the quantity $a(r_\epsilon)$ that characterizes the distinguishability: in our work, it is just a solution of a certain extremal problem, whereas in [10], it is obtained directly from the use of the respective test procedures.

The rest of the paper is organized as follows. Section 2 is concerned with the problem of finding detection boundary in a sparse Gaussian d -vectors model. In Section 3, we give a new formulation of the problem (1.3) in terms of sequence spaces. Section 4 is devoted to the description of the extremal problem that gives the distinguishability characteristics. The main results are stated in Section 5. In Section 6, we generalize the hypothesis testing problem (1.3) by considering more general alternatives. The proofs are given in Section 7.

2. Detection boundaries in a vectorial Gaussian model

Hypothesis testing problems for d -dimensional vectors, under the sparse conditions similar to the ones we use, were studied in [7, 12, 4]. Namely, let $X = (X_1, \dots, X_d)$ be a random vector of the form $X_j = v_j + \eta_j$, where $\eta_j \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$, $j = 1, \dots, d$, and

$$v_j = \xi_j a, \quad a > 0, \quad \xi_j \in \{0, 1\}, \quad K = \sum_{j=1}^d \xi_j = d^{1-b}, \quad b \in (0, 1). \quad (2.1)$$

Let $V_d(a, b) \subset \mathbb{R}^d$ be the set of all vectors $v = (v_1, \dots, v_d)$ of the form (2.1). Then, the testing problem is stated as follows: it is required to test $H_0 : v = 0$ against the alternative $H_1 : v \in V_d(a, b)$. Here the questions of interest are: what are the asymptotics for $a = a_d$ as $d \rightarrow +\infty$ for which the hypotheses H_0 and H_1 separate asymptotically? Also, what are the optimal test procedures that provide the distinguishability (or separation) of H_0 and H_1 ?

The answer to each question depends essentially on the sparsity index $b \in (0, 1)$, see [7, 12, 4]. The detection boundaries are expressed in terms of a , d and b : if $b \leq 1/2$ (moderate sparsity), then the distinguishability is impossible when $ad^{1/2-b} = o(1)$, and it is possible when $ad^{1/2-b} \rightarrow +\infty$. This is achieved by the test procedure based on a simple linear statistic $t = d^{-1/2} \sum_{i=1}^d X_i$. If $b > 1/2$ (high sparsity), then the distinguishability conditions look as follows: the distinguishability is impossible when $\limsup a/T_d < \varphi(b)$, and it is possible when $\liminf a/T_d > \varphi(b)$, where $T_d = \sqrt{\log(d)}$ and the function $\varphi(b)$, $b \in (1/2, 1)$ is defined by

$$\varphi(b) = \begin{cases} \varphi_1(b) = \sqrt{2b-1}, & 1/2 < b \leq 3/4, \\ \varphi_2(b) = \sqrt{2}(1 - \sqrt{1-b}), & 3/4 < b < 1. \end{cases} \quad (2.2)$$

Observe that the function φ is positive, continuous, and increasing in $b \in (0, 1]$.

The test procedure that provides distinguishability in the high-sparsity case is based on the Higher-Criticism statistics introduced in [4]. It is defined as $L_d = \max_{s > s_0} L_d(s)$, for any $s_0 > 0$, with

$$L_d(s) = \frac{1}{\sqrt{d\Phi(s)\Phi(-s)}} \sum_{i=1}^d (\mathbb{I}_{(X_i > s)} - \Phi(-s)), \quad (2.3)$$

where, here and later, Φ stands for the standard Gaussian cumulative distribution function. Note that it suffices to take the maximum of L_d over a discrete grid of the form $s_l = u_l T_d$, $u_l = \delta_d l$, $l = 1, \dots, L$, such that $u_L \leq \sqrt{2}$ and $\delta_d = o(1)$ is small enough.

3. Transformation of the statistical testing problem

Consider the tensor structure of the space $L_2([0, 1]^d)$, i.e., $L_2([0, 1]^d) = L_2([0, 1]) \otimes \dots \otimes L_2([0, 1])$. Then, the corresponding orthonormal basis $(\phi_l^d)_{l \in \mathbb{Z}^d}$ of $L_2([0, 1]^d)$ has the form

$$\tilde{\phi}_l^d(t) = \prod_{j=1}^d \phi_{l_j}^1(t_j), \quad t = (t_1, \dots, t_d) \in [0, 1]^d, \quad l = (l_1, \dots, l_d) \in \mathbb{Z}^d,$$

where $(\phi_k^1)_{k \in \mathbf{Z}}$ is an orthonormal basis of $L_2([0, 1])$. It is assumed that $\phi_0^1 = 1$. For any $(j, k) \in \{1, \dots, d\} \times \mathbf{Z}$, let us define $\bar{\phi}_{j,k}^d$ as

$$\bar{\phi}_{j,k}^d(t) = \tilde{\phi}_l^d(t) = \phi_k^1(t_j), \quad l = (0, \dots, k, 0, \dots, 0),$$

where k is the j -th component of l . Observe that $\bar{\phi}_{j,0}^d = 1$. Using the orthonormal system $(\bar{\phi}_{j,k}^d)_{(j,k) \in \{1, \dots, d\} \times \mathbf{Z}}$, consider the statistics $(x_j)_{1 \leq j \leq d} = \{x_{j,k}; k \in \mathbf{Z}\}_{1 \leq j \leq d}$ defined by

$$\begin{aligned} x_{j,k} &= \int_{[0,1]^d} \bar{\phi}_{j,k}^d(t) dX(t) \\ &= \xi_j \int_{[0,1]} \phi_k^1(t_j) f_j(t_j) dt_j + \epsilon \eta_{j,k} \\ &= \xi_j \theta_{j,k} + \epsilon \eta_{j,k}, \end{aligned} \tag{3.1}$$

where the random variables $\eta_{j,k} = \int_{[0,1]^d} \bar{\phi}_{j,k}^d(t) dW(t)$ are i.i.d. real standard Gaussian random variables and $\theta_{j,k} = \int_{[0,1]} \phi_k^1(t_j) f_j(t_j) dt_j$. Set $\theta_j = (\theta_{j,k})_{k \in \mathbf{Z}}$ and $\boldsymbol{\theta} = (\theta_j)_{1 \leq j \leq d}$.

Thanks to the periodic constraints, we may consider $(\phi_k^1)_{k \in \mathbf{Z}}$ as the standard Fourier basis. Then the Sobolev semi-norm of f_j can be expressed in terms of its Fourier coefficients as follows: $\|f_j\|_2^{(\tau)} = ((2\pi)^{2\tau} \sum_{k \in \mathbf{Z}} |k|^{2\tau} \theta_{j,k}^2)^{1/2}$. Therefore, the functional class $\mathcal{F}_d(\tau, r_\epsilon, b)$ can be equivalently represented as the sequence space $\Theta_d(\tau, r_\epsilon, b)$:

$$\Theta_d(\tau, r_\epsilon, b) = \{\bar{\boldsymbol{\theta}} = (\theta_1 \xi_1, \dots, \theta_d \xi_d) : \sum_{j=1}^d \xi_j = d^{1-b}; \quad \forall j \in \{1, \dots, d\}, \theta_j \in \Theta(\tau, r_\epsilon)\},$$

where

$$\Theta(\tau, r_\epsilon) = \{\theta \in l_2(\mathbf{Z}) : (2\pi)^{2\tau} \sum_{k \in \mathbf{Z}} |k|^{2\tau} \theta_k^2 \leq 1; \sum_{k \in \mathbf{Z}} \theta_k^2 \geq r_\epsilon^2\}.$$

The testing problem of interest (1.3) can be rewritten in the form

$$H_0 : \bar{\boldsymbol{\theta}} = 0 \quad \text{versus} \quad H_1 : \bar{\boldsymbol{\theta}} \in \Theta_d(\tau, r_\epsilon, b).$$

Denote by \mathcal{P}_0 and $\mathcal{P}_{\bar{\boldsymbol{\theta}}}$ the distributions under the null and alternative hypotheses, respectively. Also, denote by \mathcal{E}_0 , Var_0 , $\mathcal{E}_{\bar{\boldsymbol{\theta}}}$, and $\text{Var}_{\bar{\boldsymbol{\theta}}}$ the expectations and variances with respect to \mathcal{P}_0 and $\mathcal{P}_{\bar{\boldsymbol{\theta}}}$, respectively. The notation \mathcal{P}_{θ_j} , \mathcal{E}_{θ_j} and Var_{θ_j} also will be used: they are related to the distribution of the observations $x_j = (x_{j,k})_{k \in \mathbf{Z}}$.

For any test procedure ψ , that is, for any function measurable with respect to the observations and taking its values on the interval $[0, 1]$, let $\omega(\psi) = \mathcal{E}_0(\psi)$ be the type I error probability and let $\beta(\psi, \Theta_d(\tau, r_\epsilon, b)) = \sup_{\bar{\boldsymbol{\theta}} \in \Theta_d(\tau, r_\epsilon, b)} \mathcal{E}_{\bar{\boldsymbol{\theta}}}(1 - \psi)$ be the maximal type II error probability over the set

$\Theta_d(\tau, r_\epsilon, b)$. Also, consider the total error probability $\gamma(\psi, \Theta_d(\tau, r_\epsilon, b)) = \omega(\psi) + \beta(\psi, \Theta_d(\tau, r_\epsilon, b))$, and denote by γ or $\gamma(\Theta_d(\tau, r_\epsilon, b))$ the minimax total error probability over $\Theta_d(\tau, r_\epsilon, b)$, that is,

$$\gamma = \gamma(\Theta_d(\tau, r_\epsilon, b)) = \inf_{\psi} \gamma(\psi, \Theta_d(\tau, r_\epsilon, b)), \tag{3.2}$$

where the infimum is taken over all test procedures. One can not distinguish between H_0 and H_1 if $\gamma \rightarrow 1$, and distinguishability occurs if it exists ψ such that either $\gamma(\psi, \Theta_d(\tau, r_\epsilon, b)) \rightarrow 0$ or $\beta(\psi, \Theta_d(\tau, r_\epsilon, b)) = o(1)$ once ψ has a given asymptotic level.

The aim of this paper is to provide separation rates for the alternatives $\Theta_d(\tau, r_\epsilon, b)$ and to determine statistical procedures ψ and/or ψ_α asymptotically of level α , i.e., $\omega(\psi_\alpha) \leq \alpha + o(1)$, for which these separation rates are achieved.

By the separation rates we mean a family r_ϵ^* such that

$$\left\{ \begin{array}{l} \gamma \rightarrow 1 \\ \gamma(\psi, \Theta_d(\tau, \epsilon, b)) \rightarrow 0, \quad \text{and/or} \quad \forall \alpha \in (0, 1) \quad \beta(\psi_\alpha, \Theta_d(\tau, r_\epsilon, b)) \rightarrow 0 \end{array} \right. \quad \begin{array}{l} \text{if} \quad \frac{r_\epsilon}{r_\epsilon^*} \rightarrow 0, \\ \text{if} \quad \frac{r_\epsilon}{r_\epsilon^*} \rightarrow +\infty. \end{array}$$

By the sharp separation rates, we mean a family r_ϵ^* such that

$$\left\{ \begin{array}{l} \gamma \rightarrow 1 \\ \gamma(\psi, \Theta_d(\tau, r_\epsilon, b)) \rightarrow 0, \quad \text{and/or} \quad \forall \alpha \in (0, 1) \quad \beta(\psi_\alpha, \Theta_d(\tau, r_\epsilon, b)) \rightarrow 0 \end{array} \right. \quad \begin{array}{l} \text{if} \quad \limsup \frac{r_\epsilon}{r_\epsilon^*} < 1, \\ \text{if} \quad \liminf \frac{r_\epsilon}{r_\epsilon^*} > 1. \end{array}$$

Typically, asymptotics for models like model (1.1) are given as $\epsilon \rightarrow 0$. However, we are mainly interested in high-dimensional settings when $d \rightarrow +\infty$. Therefore, here and later, asymptotics and symbols o , O , \sim and \asymp are used when $\epsilon \rightarrow 0$ and $d \rightarrow +\infty$, except for the cases when it is explicitly specified, say, o_d is used when $d \rightarrow +\infty$. The notation $A \triangleq B$ means that we use notation A for quantity B .

4. Extremal problem

In this section, we explain what is the quantity $a(r_\epsilon)$ that corresponds to the energy of a signal in the vectorial case. Only in this section, we assume that the observations have the form $x_k = \theta_k + \epsilon \eta_k$ for $k \in \mathbf{Z}$, where the η_k 's are i.i.d. real standard Gaussian random variables. The quantity $a(r_\epsilon)$ denotes the solution of the extremal problem

$$a^2(r_\epsilon) = \frac{1}{2\epsilon^4} \inf_{\theta \in l_2(\mathbf{Z})} \sum_{k \in \mathbf{Z}} \theta_k^4 \text{ subject to } \left\{ \begin{array}{l} (2\pi)^{2\tau} \sum_{k \in \mathbf{Z}} |k|^{2\tau} \theta_k^2 \leq 1 \\ \sum_{k \in \mathbf{Z}} \theta_k^2 \geq r_\epsilon^2 \end{array} \right. \quad (4.1)$$

and characterizes distinguishability in the minimax detection problem for one-variable functions lying in \tilde{S}_τ and separated from the null hypothesis in L_2 by positive values r_ϵ , i.e., for $t \in [0, 1]$, $f(t) = \sum_{k \in \mathbf{Z}} \theta_k \phi_k^1(t)$ with $f \in \tilde{S}_\tau$ and $\|f\|_2 \geq r_\epsilon$.

Namely, if $a(r_\epsilon) \rightarrow 0$ then the minimax total error probability $\gamma(\Theta(\tau, r_\epsilon)) \rightarrow 1$, and if $a(r_\epsilon) \rightarrow +\infty$, then $\gamma(\Theta(\tau, r_\epsilon)) \rightarrow 0$.

Furthermore, let $\theta^* \triangleq \theta^*(r_\epsilon)$ be a sequence in $l_2(\mathbf{Z})$ that provides solution to the extremal problem (4.1). Set

$$w_k(r_\epsilon) = \frac{1}{2} \frac{(\theta_k^*(r_\epsilon))^2}{a(r_\epsilon)\epsilon^2}, \quad k \in \mathbf{Z}. \quad (4.2)$$

Suppose that

$$a(r_\epsilon) \asymp 1, \quad \sup_{k \in \mathbf{Z}} w_k(r_\epsilon) = o(1). \quad (4.3)$$

Then, we get the sharp asymptotics

$$\gamma(\Theta(\tau, r_\epsilon)) = 2\Phi(-a(r_\epsilon)/2) + o(1).$$

For the reader's convenience, we give a sketch of the proofs of these results. The proofs are based on the methods and results of Sections 3.1, 3.3, 4.3 in [13]. In the vectorial case in hand, we also describe the structure of asymptotically minimax tests.

In order to obtain lower bounds, we consider the Bayesian hypothesis testing problem with the product prior distribution on θ , using the symmetric two-point factors: $\pi = \prod_{k \in \mathbf{Z}} \pi_k$, $\pi_k =$

$\frac{1}{2}(\delta_{-\theta_k} + \delta_{\theta_k})$ for $\theta \in \Theta(\tau, r_\epsilon)$, and δ is the Dirac mass. Let \mathbb{P}_π be the mixture of measures \mathbb{P}_θ

over π . Observe that

$$\frac{d\mathbb{P}_\pi}{d\mathbb{P}_0}((x_k)_{k \in \mathbf{Z}}) = \prod_{k \in \mathbf{Z}} \frac{d\mathbb{P}_{\pi_k}}{d\mathbb{P}_0}(x_k) = \prod_{k \in \mathbf{Z}} \exp(-\theta_k^2/2\epsilon^2) \cosh(x_k \theta_k / \epsilon^2).$$

For the sake of simplicity, denote $\frac{d\mathbb{P}_\pi}{d\mathbb{P}_0} \triangleq \frac{d\mathbb{P}_\pi}{d\mathbb{P}_0}((x_k)_{k \in \mathbf{Z}})$. Since $\pi(\Theta(\tau, r_\epsilon)) = 1$, we have, see Proposition 2.12 in [13],

$$\gamma(\Theta(\tau, r_\epsilon)) \geq 1 - \frac{1}{2} \mathbb{E}_0 |d\mathbb{P}_\pi/d\mathbb{P}_0 - 1| \geq 1 - \frac{1}{2} (\mathbb{E}_0 (d\mathbb{P}_\pi/d\mathbb{P}_0 - 1)^2)^{1/2} = 1 - \frac{1}{2} ((\mathbb{E}_0 (d\mathbb{P}_\pi/d\mathbb{P}_0)^2) - 1)^{1/2}.$$

This yields $\gamma(\Theta(\tau, r_\epsilon)) \rightarrow 1$ as soon as $\mathbb{E}_0 (d\mathbb{P}_\pi/d\mathbb{P}_0)^2 \rightarrow 1$. Simple calculations and the inequality $\cosh(x) \leq \exp(x^2/2)$ give

$$\mathbb{E}_0 (d\mathbb{P}_\pi/d\mathbb{P}_0)^2 = \prod_{k \in \mathbf{Z}} \mathbb{E}_0 (d\mathbb{P}_{\pi_k}/d\mathbb{P}_0)^2 = \prod_{k \in \mathbf{Z}} \cosh((\theta_k/\epsilon)^2) \leq \exp\left(\frac{1}{2\epsilon^4} \sum_{k \in \mathbf{Z}} \theta_k^4\right).$$

Therefore, providing the "asymptotically least favorable prior" of the type under consideration leads to the problem (4.1).

Under assumption (4.3), taking the prior based on the extremal sequence in the problem (4.1), one can show that the Bayesian log-likelihood ratio is asymptotically Gaussian:

$$\log(d\mathbb{P}_\pi/d\mathbb{P}_0) = \sum_{k \in \mathbf{Z}} \left(-\frac{(\theta_k^*)^2}{2\epsilon^2} + \log(\cosh(x_k \theta_k^* / \epsilon^2)) \right) = -a^2(r_\epsilon)/2 + a(r_\epsilon)\eta_\epsilon + \rho_\epsilon,$$

where $\eta_\epsilon \rightarrow \eta \sim \mathcal{N}(0, 1)$ and $\rho_\epsilon \rightarrow 0$ in \mathbb{P}_0 -probability. The proof is based on Taylor's expansion, see Section 4.3.1 of [13]. This yields the sharp lower bounds.

In order to obtain upper bounds, take a sequence $q = (q_k)_{k \in \mathbf{Z}}$ such that $q_k \geq 0$, $\sum_k q_k^2 = 1/2$, and consider t_q , a centered and normalized (under \mathbb{P}_0) statistic of a weighted χ^2 -type:

$$t_q = \sum_{k \in \mathbf{Z}} q_k \left(\left(\frac{x_k}{\epsilon} \right)^2 - 1 \right).$$

Consider also the test procedures $\psi_{H,q} = \mathbb{I}_{t_q > H}$. Observe that $\mathbb{E}_0 t_q = 0$, $\text{Var}_0 t_q = 1$, and t_q are asymptotically standard Gaussian under \mathbb{P}_0 . These observations imply $w(\psi_{H,q}) = \Phi(-H) + o(1)$. Denote by $\kappa(\theta, q)$ and $\kappa(q)$ the following functions:

$$\kappa(\theta, q) = \sum_{k \in \mathbf{Z}} q_k \theta_k^2, \quad \kappa(q) = \kappa(\Theta(\tau, r_\epsilon), q) = \inf_{\theta \in \Theta(\tau, r_\epsilon)} \kappa(\theta, q). \quad (4.4)$$

Then,

$$\mathbb{E}_\theta t_q = \epsilon^{-2} \kappa(\theta, q), \quad \text{Var}_\theta t_q = 1 + 4\epsilon^{-2} \sum_k q_k^2 \theta_k^2 = 1 + O((\max_k q_k) \mathbb{E}_\theta t_q),$$

and hence, by Chebyshev's inequality, $\beta(\psi_{H,q}, \Theta(\tau, r_\epsilon)) \rightarrow 0$ when $\epsilon^{-2} \kappa(q) \rightarrow +\infty$ and $H \leq c\epsilon^{-2} \kappa(q)$, $c \in (0, 1)$. Under assumption (4.3), one can check that the statistic $\hat{t}_q = t_q - \mathbb{E}_\theta t_q$ is asymptotically standard Gaussian under \mathbb{P}_θ such that $\mathbb{E}_\theta t_q = O(1)$. Therefore

$$\beta(\psi_{H,q}, \Theta(\tau, r_\epsilon)) \leq \Phi(H - \epsilon^{-2} \kappa(q)) + o(1).$$

In order to determine "asymptotically the best sequence" $(q_k)_{k \in \mathbf{Z}}$, it suffices to find a solution of the following maximin problem:

$$\tilde{a}(r_\epsilon) = \epsilon^{-2} \sup_{\sum_k q_k^2 = 1/2, q_k \geq 0} \kappa(q). \quad (4.5)$$

First, we change the variables for $v = (v_k)_{k \in \mathbf{Z}}$ and $(p_k)_{k \in \mathbf{Z}}$, where $v_k = \theta_k^2/\sqrt{2}$, $p_k = \sqrt{2}q_k$. Then, by convexity of the set

$$V^+ = \{v \in l_1(\mathbf{Z}) : v_k \geq 0; (2\pi)^{2\tau} \sum_{k \in \mathbf{Z}} k^{2\tau} v_k \leq 2^{-1/2}; \sum_{k \in \mathbf{Z}} v_k \geq 2^{-1/2} r_\epsilon^2\}, \quad (4.6)$$

and using the minimax theorem, we get

$$\begin{aligned} \tilde{a}(r_\epsilon) &= \epsilon^{-2} \sup_{\sum_k p_k^2 = 1, p_k \geq 0} \inf_{v \in V^+} \sum_k p_k v_k = \epsilon^{-2} \sup_{\sum_k p_k^2 \leq 1, p_k \geq 0} \inf_{v \in V^+} \sum_k p_k v_k \\ &= \epsilon^{-2} \inf_{v \in V^+} \sup_{\sum_k p_k^2 \leq 1, p_k \geq 0} \sum_k p_k v_k = \epsilon^{-2} \inf_{v \in V^+} \left(\sum_k v_k^2 \right)^{1/2} \\ &= \frac{1}{\sqrt{2} \epsilon^2} \inf_{\theta \in \Theta(\tau, r_\epsilon)} \left(\sum_k \theta_k^4 \right)^{1/2} = a(r_\epsilon). \end{aligned}$$

Thus, asymptotically the best sequence $(q_k)_{k \in \mathbf{Z}}$ is the sequence $w(r_\epsilon) \triangleq (w_k(r_\epsilon))_{k \in \mathbf{Z}}$ of the form (4.2), and the value of the problem (4.5) coincides with the value of the problem (4.1). Setting $H = a(r_\epsilon)/2$, we get the upper bounds and the structure of asymptotically minimax tests.

Note that the above evaluations entail (see also Proposition 4.1 in [13]) that

$$\inf_{\theta \in \Theta(\tau, r_\epsilon)} \frac{1}{\epsilon^2} \kappa(\theta, w(r_\epsilon)) \geq a(r_\epsilon). \quad (4.7)$$

Moreover if $(\sum_{k \in \mathbf{Z}} \theta_k^2)^{1/2}$ is larger than r_ϵ , then $\kappa(\theta, w(r_\epsilon))$ becomes rather large. Namely, let us denote

$$\kappa(r_\epsilon, B) = \inf_{\theta \in \Theta(\tau, Br_\epsilon)} \kappa(\theta, w(r_\epsilon)), \quad B > 0$$

Proposition 4.1. *Let $B \geq 1$, then*

$$\frac{1}{\epsilon^2} \kappa(r_\epsilon, B) \geq B^2 a(r_\epsilon).$$

Proof of Proposition 4.1.

Set $\Theta(\tau, A, r_\epsilon) = \{\theta \in l_2(\mathbf{Z}) : (2\pi)^{2\tau} \sum_{k \in \mathbf{Z}} |k|^{2\tau} \theta_k^2 \leq A^2, \sum_{k \in \mathbf{Z}} \theta_k^2 \geq r_\epsilon^2\}$, $A > 0$. Since $\Theta(\tau, Br_\epsilon) \subset \Theta(\tau, B, Br_\epsilon)$, we have

$$\inf_{\theta \in \Theta(\tau, Br_\epsilon)} \kappa(\theta, w(r_\epsilon)) \geq \inf_{\theta \in \Theta(\tau, B, Br_\epsilon)} \kappa(\theta, w(r_\epsilon)) = B^2 \inf_{\theta \in \Theta(\tau, r_\epsilon)} \kappa(\theta, w(r_\epsilon)) \geq B^2 \epsilon^2 a(r_\epsilon),$$

where the last inequality follows from (4.7). This completes the proof.

The solution of the extremal problem (4.1) is obtained in Ingster and Suslina [13], Section 4.3. Adapting the derivations on pages 146–147 of Section 4.3.2. in [13] to our case, we set $c_3 = \frac{1}{4\tau} B(a, b)$, $c_2 = \frac{1}{4\tau} B(b, c)$ and $c_0 = \frac{1}{8\tau} B(a, d)$, where $B(\cdot, \cdot)$ is the Euler Beta function, $a = \frac{1}{2\tau}$, $b = 1 + \frac{1}{2\tau}$, $c = 2$ and $d = 3$.

Lemma 4.1. *The solution of the extremal problem (4.1) is given by*

$$a(r_\epsilon) \sim (c_1(\tau))^{1/2} r_\epsilon^{2+1/(2\tau)} \epsilon^{-2} \quad \text{as } r_\epsilon \rightarrow 0, \quad (4.8)$$

$$\text{where } c_1(\tau) = c_0 \pi c_2^{-2} \left(\frac{c_2}{c_3} \right)^{(4\tau+1)/2\tau} \text{ is a positive constant.} \quad (4.9)$$

Remark 4.1. *One must note that $r_\epsilon \rightarrow 0$ is the only condition we need to obtain the asymptotic solution of (4.1). In particular, it is not required that $\epsilon \rightarrow 0$ and Lemma 4.1 is valid whatever the value of $\epsilon > 0$ is.*

Sketch of proof of Lemma 4.1.

Following Chapter 4 in [13], observe that by setting $v_k = \theta_k^2/\sqrt{2}$ for all $k \in \mathbf{Z}$, one can transform the minimization problem under constraints (4.1) into the following one:

$$v_\epsilon^2 = \inf_{(v_k)_{k \in \mathbf{Z}} \in V^+} \sum_{k \in \mathbf{Z}} v_k^2,$$

where V^+ is defined by equation (4.6). The space $l_1^+(\mathbf{Z})$ contains non-negative sequences lying in $l_1(\mathbf{Z})$. Note that $v_\epsilon^2 = \epsilon^4 a^2(r_\epsilon)$. The convexity of the set V^+ assures the uniqueness of v_ϵ^2 . In order to determine the solution, rewrite as in Section 4.3. in [13] the sequence $(v_k)_{k \in \mathbf{Z}}$ as follows: $v_k = v_0 \zeta(k/m)$, where $\zeta(y) = (1 - |y|^{2\tau}) \mathbf{1}_{(|y| \leq 1)}$ and $m > 0$. By using the Lagrange multipliers rule, it is possible to obtain the following relations, as $r_\epsilon \rightarrow 0$ and $m \rightarrow +\infty$:

$$c_3 v_0 m \sim 2^{-1/2} r_\epsilon^2, \quad v_\epsilon^2 \sim c_0 v_0^2 m, \quad c_2 v_0 m^{2\tau+1} \sim 2^{-1/2} (2\pi)^{-2\tau}, \quad (4.10)$$

which entail the existence of v_ϵ^2 satisfying $v_\epsilon^2 \sim c_1(\tau) r_\epsilon^{4+1/\tau}$, and thus $a^2(r_\epsilon) \sim c_1(\tau) \epsilon^{-4} r_\epsilon^{4+1/\tau}$.

If $r_\epsilon \rightarrow 0$, then the first and second relations in (4.10) entail that

$$v_0 \asymp v_\epsilon^2 r_\epsilon^{-2} \asymp r_\epsilon^{2+1/\tau}, \quad (4.11)$$

which implies that $m \rightarrow +\infty$ since the third relation in (4.10) yields $m \asymp v_0^{-1/(2\tau+1)} \asymp r_\epsilon^{-1/\tau}$.

Remark 4.2. The form of function ζ and relation (4.11) imply that $\sup_k v_k \leq v_0 = o(1)$.

5. Main results

Depending on the values of b , we distinguish between two types of sparsity: the moderate sparsity case with $b \in (0, 1/2]$ and the high sparsity case with $b \in (1/2, 1)$. In each case, although being of different types, the “best” test procedures that achieve the separation rates are based on the χ^2 -type statistics $(t_j)_{1 \leq j \leq d}$ determined in the same way as the “best statistic” t_q of a weighted χ^2 -type in Section 4.

Let us introduce a general version of the χ^2 -type statistics of interest. For j in $\{1, \dots, d\}$, put

$$t_j = \sum_{k \in \mathbf{Z}} w_k \left(\left(\frac{x_{j,k}}{\epsilon} \right)^2 - 1 \right), \quad (5.1)$$

where $(w_k)_{k \in \mathbf{Z}}$ is the sequence of weights such that $w_k \geq 0$ for all k in \mathbf{Z} and $\sum_{k \in \mathbf{Z}} w_k^2 = \frac{1}{2}$. Set also

$$t_{j,k} = w_k \left(\left(\frac{x_{j,k}}{\epsilon} \right)^2 - 1 \right), \quad (5.2)$$

so that $t_j = \sum_{k \in \mathbf{Z}} t_{j,k}$.

Recall that $T_d = \sqrt{\log d}$ (see Section ??). Similarly to (2.3) and for any $u \in (0, \sqrt{2}]$, let us define the statistics $L(u)$ on which the Higher-Criticism test procedure is built:

$$L(u) = C_u \sum_{j=1}^d (\mathbf{1}_{(t_j > u T_d)} - \tilde{\Phi}_0(u T_d)), \quad (5.3)$$

where

$$\tilde{\Phi}_0(x) = \mathbb{P}_0(t_j > x) \quad (5.4)$$

$$C_u = (d \tilde{\Phi}_0(u T_d) (1 - \tilde{\Phi}_0(u T_d)))^{-1/2}. \quad (5.5)$$

Taking into account the sparsity condition, we consider a particular sequence of weights $(w_k(r_\epsilon^*))_{k \in \mathbf{Z}}$ defined by equation (4.2) with $r_\epsilon^* \triangleq r_\epsilon^*(b)$ being the separations rates depending

on b in $(0, 1)$. Then, for all $j \in \{1, \dots, d\}$, we consider the statistics $t_{j,b}$ as in (5.1) with the weight sequence $(w_k(r_\epsilon^\star))_{k \in \mathbf{Z}}$, that is,

$$t_{j,b} = \sum_{k \in \mathbf{Z}} w_k(r_\epsilon^\star) \left(\left(\frac{x_{j,k}}{\epsilon} \right)^2 - 1 \right).$$

Also, denote by t_b the normalized empirical mean of the $t_{j,b}$'s:

$$t_b = \frac{1}{\sqrt{d}} \sum_{j=1}^d t_{j,b}. \quad (5.6)$$

Similarly, replacing t_j by $t_{j,b}$, consider the statistics $L(u, b)$, $C_{u,b}$, and $\tilde{\Phi}_{0,b}$ defined by equations (5.3), (5.5) and (5.4) respectively, that is,

$$\begin{aligned} L(u, b) &= C_{u,b} \sum_{j=1}^d (\mathbb{I}_{(t_{j,b} > uT_d)} - \tilde{\Phi}_{0,b}(uT_d)), \\ C_{u,b} &= (d\tilde{\Phi}_{0,b}(uT_d)(1 - \tilde{\Phi}_{0,b}(uT_d)))^{-1/2}, \\ \tilde{\Phi}_{0,b}(x) &= \mathbb{P}_0(t_{j,b} > x). \end{aligned} \quad (5.7)$$

5.1. Moderate sparsity

In case of moderate sparsity, for any $\alpha \in (0, 1)$, consider the χ^2 -type test procedure:

$$\psi_\alpha^{\chi^2} \triangleq \psi_{\alpha,b}^{\chi^2} = \mathbb{I}_{(t_b > T_\alpha)}, \quad (5.8)$$

where t_b is defined in (5.6) and T_α is the $(1 - \alpha)$ -quantile of a real standard Gaussian random variable.

Theorem 5.1. *Assume that $r_\epsilon \rightarrow 0$ and let $a(r_\epsilon)$ be given by (4.8). Then, the following results hold true.*

- (i) *Lower bound.*
 If $a(r_\epsilon)d^{1/2-b} = o(1)$, then $\gamma \rightarrow 1$.
 If $a(r_\epsilon)d^{1/2-b} = O(1)$, then $\liminf \gamma > 0$.
- (ii) *Upper bound.* Let $r_\epsilon^\star = r_\epsilon^\star(b)$ be determined by the relation $a(r_\epsilon^\star) \asymp d^{b-1/2}$ and $\psi_\alpha^{\chi^2}$ be defined by (5.8). Then,
 Type I error: $\forall \alpha \in (0, 1)$, $\omega(\psi_\alpha^{\chi^2}) = \alpha + o(1)$.
 Type II error: if $a(r_\epsilon)d^{1/2-b} \rightarrow +\infty$, then $\beta(\psi_\alpha^{\chi^2}, \Theta_d(\tau, r_\epsilon, b)) = o(1)$.

Remark 5.1. *Note that we obtain the same detection boundaries as in the vectorial case (see Section 2): the areas of distinguishability and non-distinguishability depend on the limit of $d^{1/2-b}a(r_\epsilon)$. The condition $d^{1/2-b}a(r_\epsilon) \asymp a(r_\epsilon)/a(r_\epsilon^\star) \rightarrow +\infty$ is equivalent to $r_\epsilon/r_\epsilon^\star \rightarrow +\infty$ where by (4.8)*

$$r_\epsilon^\star \asymp (\epsilon^4 d^{2b-1})^{\tau/(4\tau+1)}. \quad (5.9)$$

In order to use Lemma 4.1, the condition $r_\epsilon \rightarrow 0$ is required. Note that the requirement $r_\epsilon^\star \rightarrow 0$ is always fulfilled for $b \in (0, 1/2)$ whatever the value of $\epsilon > 0$ is as soon as $d \rightarrow +\infty$. For $b = 1/2$, the condition $r_\epsilon^\star \rightarrow 0$ holds when $\epsilon \rightarrow 0$.

5.2. High sparsity

Let us define the Higher-Criticism type test procedure. Let $r_\epsilon^* = r_\epsilon^*(b)$ be determined by the relation $a(r_\epsilon^*) \sim \varphi(b)T_d$, where $\varphi(b)$ is given by (2.2). Set $u(b) = \min(2\varphi(b), \sqrt{2})$, i.e., $u(b) = 2\varphi(b)$ for $b \in (1/2, 3/4]$, and $u(b) = \sqrt{2}$ for $b \in (3/4, 1]$. Consider the test

$$\psi^L = \mathbb{I}_{\{\max_{1 \leq l \leq N-1} L(u_l, b_l) > H\}}, \quad u_l = u(b_l),$$

where the function L is defined in (5.7) and $(b_l)_{1 \leq l \leq N}$ consists of a regular grid on $(1/2, 1]$, that is, $b_l = 1/2 + l\delta$, where δ is a positive parameter that satisfies $\delta = o_d(1)$, $T_d\delta \rightarrow +\infty$ and $N\delta = 1/2$. This entails that $N = O_d(\delta^{-1})$ and thus $N = o_d(T_d)$. Take a positive H such that $H \sim (\log d)^C$ for some positive constant C satisfying $C > \frac{1}{4}$.

For a constant $D > \sqrt{2}$, consider also the test

$$\psi^{max} = \mathbb{I}_{\{\max_{1 \leq j \leq d} \max_{1 \leq l \leq N} t_{j, b_l} > DT_d\}}.$$

Finally, combining ψ^L and ψ^{max} , we define the test procedure

$$\psi^{HC} = \psi^L \psi^{max}, \quad (5.10)$$

that rejects H_0 if both ψ^L and ψ^{max} reject H_0 .

For the high sparsity case, not only separation rates but also sharp asymptotics are obtained; two ranges of b should be distinguished: the range of b in $(1/2, 3/4]$, called the intermediate sparsity case, and the range of b in $(3/4, 1]$, called the highest sparsity case.

Theorem 5.2. *Assume that $r_\epsilon \rightarrow 0$ and that $\log d = o(\epsilon^{-2/(2\tau+1)})$. Let $a(r_\epsilon)$ be given by (4.8) and let φ be given by (2.2).*

- (i) *Lower bound. If $\limsup a(r_\epsilon)/T_d < \varphi(b)$, then $\liminf \gamma \rightarrow 1$.*
- (ii) *Upper bound: errors of ψ^{HC} defined by (5.10).*
 - *Type I error: $\omega(\psi^{HC}) = o(1)$.*
 - *Type II error: if $\liminf a(r_\epsilon)/T_d > \varphi(b)$, then $\beta(\psi^{HC}, \Theta_d(\tau, r_\epsilon, b)) = o(1)$.*

Remark 5.2. • *Set $a(r_\epsilon^*) = T_d\varphi(b)$. In our sparse functional framework, the distinguishability conditions are the same as for a d -dimensional sparse vector (see, e.g., [12]), with the only difference that in our case the assumption $\log d = o(\epsilon^{-2/(2\tau+1)})$ is required. Under this assumption, the result of Theorem 5.2 means that distinguishability is impossible if $\limsup a(r_\epsilon)/a(r_\epsilon^*) < 1$ and it is possible if $\liminf a(r_\epsilon)/a(r_\epsilon^*) > 1$. Due to (4.8), these conditions provide sharp separation rates since they are equivalent to $\limsup r_\epsilon/r_\epsilon^* < 1$ and $\liminf r_\epsilon/r_\epsilon^* > 1$, respectively, where*

$$r_\epsilon^* \sim (\epsilon^4 T_d^2 (c_1(\tau))^{-1} \varphi^2(b))^{\tau/(4\tau+1)}, \quad (5.11)$$

and $c_1(\tau)$ is defined by (4.9). Note that the condition $r_\epsilon^* \rightarrow 0$ is fulfilled under the assumption $\log d = o(\epsilon^{-2/(2\tau+1)})$.

The values r_ϵ^* mark the border between the areas of distinguishability and non-distinguishability. Indeed, for $r_\epsilon \rightarrow 0$ such that $\limsup r_\epsilon/r_\epsilon^* < 1$, the alternatives separated from the null hypothesis by r_ϵ are not distinguishable and, on the other side, for $r_\epsilon \rightarrow 0$ such that $\liminf r_\epsilon/r_\epsilon^* > 1$, the alternatives separated from the **null hypothesis** by r_ϵ are distinguishable.

- *Actually, the assumption $\log d = o(\epsilon^{-2/(2\tau+1)})$ is equivalent to*

$$(r_\epsilon^*)^{1/(2\tau)} T_d = o(1), \quad (5.12)$$

which is required when dealing with the asymptotic behavior of the tail distribution of $t_{j,b}$ (see Lemma 7.1) since $T_d \sup_k w_k(r_\epsilon^*) \leq (r_\epsilon^*)^{1/(2\tau)} T_d$. Relation (5.12) follows from the relations in (4.10). Concerning the lower bound, condition (5.12) is necessary when we evaluate the second moment of the Bayesian likelihood ratio under the null hypothesis.

- Note that the condition $\log d = o(\epsilon^{-2/(2\tau+1)})$ is essential for $b \in (1/2, 1)$. Namely, it follows from Theorem 2 in [10] that if $\liminf (\log d \epsilon^{2/(2\tau+1)}) > 0$, then the separation rates are of the form $r_\epsilon^* = \epsilon \sqrt{\log d}$ for any $b \in (1/2, 1)$. Observe that if $\log d \geq c\epsilon^{-2}$ for some $c > 0$, then the separation rates are bounded away from zero, i.e., it is impossible to detect functions lying in $\Theta_d(\tau, r_\epsilon, b)$ with small enough $r_\epsilon > 0$.

Remark 5.3. ADAPTATION.

In the high sparsity case, a family of test procedures ψ^{HC} provides the distinguishability for all $b \in (1/2, 1)$. Moreover, it follows from the proofs that our result is uniform over $b \in (1/2 + \rho, 1 - \rho)$ for any $\rho \in (0, 1/4)$, i.e., the results are adaptive over $b \in (1/2 + \rho, 1 - \rho)$ for any $\rho \in (0, 1/4)$, without a loss in separation rates.

For the moderate sparsity case, it is worth noting that the family of test procedures $\psi_\alpha^{\chi^2} = \psi_{\alpha, b}^{\chi^2}$ depends on $b \in (0, 1/2]$ since the sequence of weights $w(r_\epsilon^*(b))$ does. It is shown in Theorem 3 of [10] that “adaptive” separation rates for unknown $b \in (0, 1/2)$ are of the form $r_\epsilon^* \asymp (\epsilon^4 d^{2b-1} \log \log d)^{\tau/(4\tau+1)}$, i.e., the adaptive case leads to an unavoidable $\log \log$ -loss in separation rates compared to non-adaptive setting. Using the Bonferroni method, it is possible to prove that the test procedures based on a grid of tests of the form $\psi_{\alpha_d, b_l}^{\chi^2}$ are adaptive rate-optimal test procedures. Since this result is similar to the one stated in [10], we omit it.

6. Extended problem

In this section, we generalize the hypothesis testing problem stated in (1.3) to more general alternatives. The null hypothesis H_0 is still characterized by some constant $const_0$ and, as in (1.3), under the alternative, the signal function f is, up to some constant, equal to f^1 , i.e., $f = const_1 + f^1$. The additive sparse structure on f^1 is still assumed, i.e., $f^1 \in \mathcal{F}_{d,b}$, as well as every component f_j^1 is assumed 1-periodic and orthogonal to a constant (recall that for any $t \in [0, 1]^d$ $f^1(t) = \sum_{j=1}^d \xi_j f_j^1(t_j)$ where $\xi_j \in \{0, 1\}$ and $t_j \in [0, 1]$ for any $j \in \{1, \dots, d\}$). We then denote by $\tilde{\mathcal{F}}_{d,b}$ the set of signal functions in $\mathcal{F}_{d,b}$ whose components are 1-periodic and orthogonal to a constant. Rather than imposing smoothness constraints component-wise, we now study the alternative classes for which the smoothness and separation conditions are expressed in terms of the whole signal function f^1 . In other words, the main difference between the extended and initial detection problems is that the distinguishability problem is studied with respect to a global signal.

Then, given the alternatives that include signal functions f as in (1.3), where f^1 belongs to the functional class $\mathcal{F}_d^{ext}(\tau, L, r_\epsilon, b)$, the testing problem of interest is stated as follows:

$$H_0 : f = const_0 \quad \text{versus} \quad H_1 : f = const_1 + f^1, \quad f^1 \in \mathcal{F}_d^{ext}(\tau, L, r_\epsilon, b), \quad (6.1)$$

where

$$\mathcal{F}_d^{ext}(\tau, L, r_\epsilon, b) = \left\{ f^1 \in \tilde{\mathcal{F}}_{d,b} : \|f^1\|_2 \geq r_\epsilon, \|f^1\|_2^{(\tau)} \leq L \right\},$$

in which $(\|f^1\|_2^{(\tau)})^2 = \sum_{j=1}^d \xi_j (\|f_j^1\|_2^{(\tau)})^2$. Due to the periodic constraint, we consider the standard Fourier basis. This allows to express the semi-norm $\|\cdot\|_2^{(\tau)}$ in terms of Fourier coefficients. As in Section 3, we then transform the functional space $\mathcal{F}_d^{ext}(\tau, r_\epsilon, L, b)$ to the sequence space $\Theta_d^{ext}(\tau, L, r_\epsilon, b)$, which consists of sequences $\bar{\theta} = (\xi_j \theta_{j,k})_{j,k}$ such that

$$\begin{aligned} \sum_{j=1}^d \xi_j &= d^{1-b} = K, \\ \sum_{j=1}^d \xi_j (2\pi)^{2\tau} \sum_{k \in \mathbf{Z}} |k|^{2\tau} \theta_{j,k}^2 &\leq L^2, \\ \sum_{j=1}^d \xi_j \sum_{k \in \mathbf{Z}} \theta_{j,k}^2 &\geq r_\epsilon^2. \end{aligned}$$

Note that if $L^2 = K$ and $\tilde{r}_\epsilon^2 = Kr_\epsilon^2$, then we have

$$\Theta_d^{ext}(\tau, L, \tilde{r}_\epsilon, b) \supset \Theta_d(\tau, r_\epsilon, b).$$

This implies that the results on the lower bound continue to hold for $\Theta_d^{ext}(\tau, L, \tilde{r}_\epsilon, b)$ with the separation rates $(\tilde{r}_\epsilon^*)^2 = K(r_\epsilon^*)^2$, where r_ϵ^* is defined by either (5.9) or (5.11) depending on the values of b . Here, the quantity of interest is $\tilde{a}(r_\epsilon)$, the solution of the following extremal problem:

$$\tilde{a}^2(r_\epsilon) = \frac{1}{2\epsilon^4} \inf_{\bar{\theta} \in l_2} \sum_{j=1}^d \xi_j \sum_{k \in \mathbf{Z}} \theta_{j,k}^4 \text{ subject to } \begin{cases} \sum_{j=1}^d \xi_j = d^{1-b} = K \\ \sum_{j=1}^d \xi_j (2\pi)^{2\tau} \sum_{k \in \mathbf{Z}} |k|^{2\tau} \theta_{j,k}^2 \leq K \\ \sum_{j=1}^d \xi_j \sum_{k \in \mathbf{Z}} \theta_{j,k}^2 \geq Kr_\epsilon^2 \end{cases} \quad (6.2)$$

As follows from Section 4.3 in [13], the solution of the extremal problem (6.2) is given by

$$\tilde{a}(r_\epsilon) \sim (c_1(\tau))^{1/2} Kr_\epsilon^{2+1/(2\tau)} \epsilon^{-2} \text{ as } r_\epsilon \rightarrow 0,$$

where $c_1(\tau)$ is defined in (4.9). That is, $\tilde{a}(r_\epsilon) = Ka(r_\epsilon)$, where $a(r_\epsilon)$ is the solution (4.8) of the extremal problem (4.1).

Remark 6.1. Consider the function κ defined by (4.4), for which the sequence of weights $w(r_\epsilon) = (w_k(r_\epsilon))_k$ is defined as in (4.2). Then we obtain from (4.7) that

$$\inf_{\bar{\theta} \in \Theta_d^{ext}(\tau, K^{1/2}, K^{1/2}r_\epsilon, b)} \frac{1}{\epsilon^2} \sum_{j=1}^d \xi_j \kappa(\theta_j, w(r_\epsilon)) \geq \tilde{a}(r_\epsilon) = Ka(r_\epsilon), \quad (6.3)$$

and similarly to Proposition 4.1 for any $D \geq 1$,

$$\inf_{\bar{\theta} \in \Theta_d^{ext}(\tau, K^{1/2}, DK^{1/2}r_\epsilon, b)} \frac{1}{\epsilon^2} \sum_{j=1}^d \xi_j \kappa(\theta_j, w(r_\epsilon)) \geq D^2 \tilde{a}(r_\epsilon) = D^2 Ka(r_\epsilon). \quad (6.4)$$

Now, as in Section 3, with the use of the orthonormal system, instead of considering the random process $X(t)$ defined in model (1.1), we observe a family of random sequences $(x_{j,k})_{k \in \mathbf{Z}, j \in \{1, \dots, d\}}$ defined by (3.1). Finally, the remained question is: do the families of test procedures $\psi_\alpha^{\chi^2}$ given by (5.8) and ψ^{HC} given by (5.10) provide distinguishability? The answer is affirmative and is given below. Note that it is then sufficient to study the type II error probability of these tests since their type I error probability has been already studied for the hypothesis testing problem (1.3).

Theorem 6.1. Assume that $r_\epsilon \rightarrow 0$ and let $a(r_\epsilon)$ and φ be given by (4.8) and (2.2), respectively. Then, the following results hold true.

- (i) MODERATE SPARSITY-Type II error probability of $\psi_\alpha^{\chi^2}$ defined by (5.8).
If $a(r_\epsilon)d^{1/2-b} \rightarrow +\infty$, then $\beta(\psi_\alpha^{\chi^2}, \Theta_d^{ext}(\tau, K^{1/2}, K^{1/2}r_\epsilon, b)) = o(1)$.
- (ii) HIGH SPARSITY-Type II error probability of ψ^{HC} defined by (5.10).
Assume that $\log d = o(\epsilon^{-2/(2\tau+1)})$.
If $\liminf a(r_\epsilon)/T_d > \varphi(b)$, then $\beta(\psi^{HC}, \Theta_d^{ext}(\tau, K^{1/2}, K^{1/2}r_\epsilon, b)) = o(1)$.

Remark 6.2. One should note that the detection boundaries are the same for the hypothesis testing problems (1.3) and (6.1), the initial one and its generalization.

7. Proofs

Proofs of our main results require some preliminary results that are stated below both under the null and alternative hypotheses. Specifically, we establish asymptotic tail distributions of the test statistics in hand and find their first and second moments.

7.1. Properties of test statistics

In this section, we consider the statistics t_j defined by (5.1) with any sequence of weights $w = (w_k)_{k \in \mathbf{Z}}$ such that $w_k \geq 0, \forall k \in \mathbf{Z}$ and $\sum_k w_k^2 = 1/2$. Therefore the quantities $L(u)$, $C(u)$, and $\tilde{\Phi}_0$ are those defined by (5.3), (5.5) and (5.4).

Proposition 7.1. ASYMPTOTIC TAIL DISTRIBUTION OF t_j DEFINED BY (5.1).

Assume $T \max_k w_k = o(1)$, then

$$\begin{aligned} \log \mathbb{P}_0(t_j > T) &\sim -\frac{T^2}{2} \text{ as } T \rightarrow +\infty, \\ \log \mathbb{P}_{\theta_j}(t_j > T) &\sim -\frac{(T - \mathbb{E}_{\theta_j}(t_j))^2}{2}, \text{ as } (T - \mathbb{E}_{\theta_j}(t_j)) \xrightarrow{T \rightarrow +\infty} +\infty. \end{aligned}$$

Proof of Proposition 7.1.

We consider only the distribution \mathbb{P}_{θ_j} since \mathbb{P}_0 is a particular case of \mathbb{P}_{θ_j} . The proof consists of bounding $\mathbb{P}_{\theta_j}(t_j > T)$ from above and below. This is done by using the cumulant-generating function of t_j under \mathbb{P}_{θ_j} which is defined by $\phi_{\theta_j}(h) = \log(\mathbb{E}_{\theta_j}(\exp(ht_j)))$ for any h . Let us consider only positive h and let us introduce a new family of probability measures $\mathbb{P}_{\theta_j,h}$ such that $\frac{d\mathbb{P}_{\theta_j,h}}{d\mathbb{P}_0} = \exp(ht_j) \exp(-\phi_{\theta_j}(h))$. This yields

$$\begin{aligned} \mathbb{P}_{\theta_j}(t_j > T) &= \mathbb{E}_{\theta_j,h}[\mathbb{1}_{(t_j > T)} \exp(-(ht_j - \phi_{\theta_j}(h)))] \\ &= \exp(-(hT - \phi_{\theta_j}(h))) \mathbb{E}_{\theta_j,h}[\mathbb{1}_{(t_j > T)} \exp(-h(t_j - T))]. \end{aligned} \quad (7.1)$$

Let us start with the upper bound.

Upper bound. The second term of the right-hand side of (7.1) is less than 1. Hence there is a straightforward upper bound on $\mathbb{P}_{\theta_j}(t_j > T)$:

$$\mathbb{P}_{\theta_j}(t_j > T) \leq \exp(-(hT - \phi_{\theta_j}(h))). \quad (7.2)$$

To complete this part of the proof, it remains to determine the minimum value of a positive value h on the right-hand side of (7.2). The minimum is attained for positive h such that

$$\mathbb{E}_{\theta_j,h}(t_j) = T \quad (7.3)$$

since

$$\begin{cases} (\phi_{\theta_j}(h) - hT)' &= \mathbb{E}_{\theta_j,h}(t_j) - T, \\ (\phi_{\theta_j}(h) - hT)'' &= \text{Var}_{\theta_j,h}(t_j) \geq 0, \end{cases}$$

where $(\cdot)'$ and $(\cdot)''$ denote the first and second derivatives with respect to h , respectively, and, $\mathbb{E}_{\theta_j,h}$ and $\text{Var}_{\theta_j,h}$ are the expectation and variance with respect to $\mathbb{P}_{\theta_j,h}$.

In order to find h that solves equation (7.3), we need to determine ϕ_{θ_j} . For this, set $\nu_{j,k} = \frac{\theta_{j,k}}{\epsilon}$.

Then for any positive h such that $h \rightarrow +\infty$ and $h \max_k w_k = o(1)$, we obtain

$$\begin{aligned}
\phi_{\theta_j}(h) &= \log \prod_k \mathbb{E}_{\theta_j} [\exp(hw_k((\nu_{j,k} + \eta_{j,k})^2 - 1))] \\
&= \sum_k \left\{ -hw_k + \frac{hw_k \nu_{j,k}^2}{(1 - 2hw_k)} - \frac{1}{2} \log(1 - 2hw_k) \right\} \\
&= \sum_k \left\{ -hw_k + hw_k \nu_{j,k}^2 (1 + 2hw_k + o(hw_k)) \right. \\
&\quad \left. - \frac{1}{2} (-2hw_k - \frac{(2hw_k)^2}{2} + o(h^2 w_k^2)) \right\} \\
&= \sum_k \{ hw_k \nu_{j,k}^2 (1 + o(h \max_k w_k)) + h^2 w_k^2 (2\nu_{j,k}^2 + 1) + o(h^2 w_k^2) \} \\
&= h \mathbb{E}_{\theta_j}(t_j) (1 + o(h \max_k w_k)) + \frac{h^2}{2} (1 + o(1)) + o(h^2), \tag{7.4}
\end{aligned}$$

where the last equality sign in (7.4) follows from $(T - \mathbb{E}_{\theta_j}(t_j)) \rightarrow +\infty$ and $T \max_k w_k = o(1)$ as $T \rightarrow +\infty$. Next, differentiating the right-hand side of (7.4) with respect to h yields

$$(\phi_{\theta_j}(h) - hT)' = 0 \Rightarrow h \sim T - \mathbb{E}_{\theta_j}(t_j), \text{ as } T - \mathbb{E}_{\theta_j}(t_j) \text{ goes to infinity.}$$

As $(T - \mathbb{E}_{\theta_j}(t_j)) \xrightarrow{T \rightarrow +\infty} +\infty$, this leads to the following optimal upper bound for right-hand side of (7.2):

$$\exp \left((T - \mathbb{E}_{\theta_j}(t_j)) \mathbb{E}_{\theta_j}(t_j) + \frac{(T - \mathbb{E}_{\theta_j}(t_j))^2}{2} - T(T - \mathbb{E}_{\theta_j}(t_j)) \right) \sim \exp \left(-\frac{(T - \mathbb{E}_{\theta_j}(t_j))^2}{2} \right).$$

Since by assumption $T \max_k w_k = o(1)$, the condition $h \max_k w_k = o(1)$ with $(T - \mathbb{E}_{\theta_j}(t_j))$ in place of h is fulfilled.

By assumption $T \max_k w_k = o(1)$, hence the optimal upper bound under \mathbb{P}_0 is $\exp(-\frac{T^2}{2})$ as T goes to infinity. This completes the proof of the upper bound.

Lower Bound. We are interested in obtaining a lower bound for (7.1). This is done by first considering a new family of probability distributions under which the normalized statistics t_j are proved to be asymptotically Gaussian.

For $h > 0$ satisfying equation (7.3), let us introduce the following probability measures $\mathbb{P}_{\theta_j, h, k}$:

$$\frac{d\mathbb{P}_{\theta_j, h, k}}{d\mathbb{P}_0} = \exp(ht_{j,k}) \exp(-\phi_{\theta_j, k}(h)),$$

with $t_{j,k}$ defined in (5.2), $\phi_{\theta_j, k}(h) = \log \mathbb{E}_{\theta_j, k}(\exp(ht_{j,k}))$ and where $\mathbb{E}_{\theta_j, k}$ stands for the expectation with respect to the observations $(x_{j,k})_{j,k}$ of (3.1). Denote by $\mathbb{E}_{\theta_j, h, k}$ and $\text{Var}_{\theta_j, h, k}$ the expectation and variance with respect to $\mathbb{P}_{\theta_j, h, k}$.

To establish the asymptotic normality of t_j , we will check that the Lyapunov condition is satisfied. To this end, set $\sigma_{j, h, k}^2 = \text{Var}_{\theta_j, h, k}(t_{j,k})$ and $\sigma_{j, h}^2 = \sum_k \sigma_{j, h, k}^2$.

Denote by $\phi_{\theta_j, k}^{(2)}$ and $\phi_{\theta_j, k}^{(4)}$ the second and fourth derivatives of $\phi_{\theta_j, k}$ with respect to h , respectively. Using well-known relations between moments of t_j under $\mathbb{P}_{\theta_j, h, k}$ and the successive derivatives of

$\phi_{\theta_j,k}(h)$ with respect to h , in particular, $\sigma_{j,h}^2 = \sum_k \phi_{\theta_j,k}^{(2)}$, we get

$$\begin{aligned} \frac{\sum_k \mathbb{E}_{\theta_j,h,k}(t_{j,k} - \mathbb{E}_{\theta_j,h,k}(t_{j,k}))^4}{(\sum_k \sigma_{j,h,k}^2)^2} &= \frac{3 \sum_k (\phi_{\theta_j,k}^{(2)}(h))^2 + \sum_k \phi_{\theta_j,k}^{(4)}(h)}{(\sum_k \phi_{\theta_j,k}^{(2)}(h))^2} \\ &\leq \frac{4 \max(w_k^2) \sum_k w_k^2 (1 + o(1)) + o(1)}{1} \\ &= o(1), \end{aligned}$$

where the last relation follows from $\max w_k = o(1)$ and relation (7.4), since by (7.4) we get $\sum_k \phi_{\theta_j,k}^{(4)}(h) = \phi_{\theta_j}^{(4)}(h) = o(1)$. The Lyapunov condition is then satisfied. This implies that under $\mathbb{P}_{\theta_j,h}$, $Z_{j,h} = \frac{t_j - \mathbb{E}_{\theta_j,h}(t_j)}{\sigma_{j,h}}$ is asymptotically a real standard Gaussian random variable.

Let us return to relation (7.1), where h is chosen to have $\mathbb{E}_{\theta_j,h}(t_j) = T$, and observe that

$$\mathbb{E}_{\theta_j,h}[\mathbb{1}_{(t_j > T)} \exp(-h(t_j - T))] = \mathbb{E}_{\theta_j,h}[\mathbb{1}_{(Z_{j,h} > 0)} \exp(-hZ_{j,h}\sigma_{j,h})].$$

Due to the asymptotic normality of t_j , for any $\delta > 0$,

$$\begin{aligned} \mathbb{E}_{\theta_j,h}[\mathbb{1}_{(Z_{j,h} > 0)} \exp(-hZ_{j,h}\sigma_{j,h})] &= \mathbb{E}_{\theta_j,h}[\mathbb{1}_{(Z_{j,h} \in (0,\delta))} \exp(-hZ_{j,h}\sigma_{j,h})] + \\ &\quad \mathbb{E}_{\theta_j,h}[\mathbb{1}_{(Z_{j,h} > \delta)} \exp(-hZ_{j,h}\sigma_{j,h})] \\ &> (\mathbb{P}_{\theta_j,h}(Z_{j,h} \in (0,\delta)) + o(1)) \exp(-h\delta\sigma_{j,h}). \end{aligned} \quad (7.5)$$

By choosing $\delta = o(h)$ in relation (7.5) implies that

$$\log(\mathbb{P}_{\theta_j}(t_j > T)) \geq \phi_{\theta_j}(h) - hT - o(h^2). \quad (7.6)$$

Up to $o(h^2)$, the right-hand side of (7.6) corresponds to the argument of the exponential function on the right-hand side of (7.2). This entails that the right-hand side of (7.6) is equivalent to $-\frac{(T - \mathbb{E}_{\theta_j}(t_j))^2}{2}$. This completes the proof of the lower bound, and thus Proposition 7.1 is proved.

Lemma 7.1. • (i) Expectation and variance of t_j defined by (5.1).

$$\mathbb{E}_{\theta_j}(t_j) = \xi_j \epsilon^{-2} \kappa(\theta_j, w), \quad (7.7)$$

$$\text{Var}_{\theta_j}(t_j) = 1 + O((\max_{k \in \mathbf{Z}} w_k) \mathbb{E}_{\theta_j}(t_j)). \quad (7.8)$$

- (ii) Expectation and variance of $L(u)$ defined by (5.3). Assume that $T_d \max w_k = o(1)$ and consider any $\bar{\theta} = (\xi_1 \theta_1, \dots, \xi_d \theta_d)$ such that $\sum_{j=1}^d \xi_j = d^{1-b}$. Moreover, assume that for all nonzero ξ_j , $\mathbb{E}_{\theta_j}(t_j) \geq cT_d$, with some positive c , and $\max_{j: \xi_j=1} \mathbb{E}_{\theta_j}(t_j) = O(T_d)$. Then, for all $u \in (0, \sqrt{2}]$,

$$\begin{aligned} \mathbb{E}_{\bar{\theta}}(L(u)) &\geq d^{\frac{1}{2}-b+(\frac{u^2}{4}-\frac{((u-\epsilon)_+)^2}{2})(1+o(1))}(1+o(1)), \\ \text{Var}_{\bar{\theta}}(L(u)) &= o(d^\eta \mathbb{E}_{\bar{\theta}}(L(u))) + o(1), \quad \eta = o(1), \end{aligned}$$

where $x_+ = \max(0, x)$.

Remark 7.1. Under \mathbb{P}_0 , the statistics t_j and $L(u)$ have zero mean and unit variance. Moreover, under \mathbb{P}_0 and the assumption $\max_k w_k = o(1)$, the statistics t_j are asymptotically standard Gaussian. Under \mathbb{P}_{θ_j} , the statistics $t_j - \mathbb{E}_{\theta_j} t_j$ are asymptotically standard Gaussian if $\max_k w_k \mathbb{E}_{\theta_j} t_j = o(1)$, see Lemma 3.1 in [13].

Proof of Lemma 7.1.

- (i) Recall that $\sum_k w_k^2 = 1/2$. For each index j satisfying $\xi_j = 1$, the random variable $(\frac{x_{j,k}}{\epsilon})^2$ is a

\mathbb{P}_{θ_j} -noncentral $\chi^2(1, \theta_{j,k}^2 \epsilon^{-2})$. From this relation (7.7) is easily obtained. Relation (7.8) is deduced from the following calculations:

$$\begin{aligned} \text{Var}_{\theta_j}(t_j) &= \sum_{k \in \mathbf{Z}} w_k^2 (2 + 4\epsilon^{-2} \xi_j \theta_{j,k}^2) \\ &= 1 + \sum_{k \in \mathbf{Z}} w_k^2 4\epsilon^{-2} \xi_j \theta_{j,k}^2 \\ &= 1 + O(\max_{k \in \mathbf{Z}} w_k \epsilon^{-2} \xi_j \kappa(\theta_j, w)) \\ &= 1 + O(\max_{k \in \mathbf{Z}} w_k \mathbb{E}_{\theta_j}(t_j)). \end{aligned}$$

(ii) For any $u \in (0, \sqrt{2}]$, as $T_d \rightarrow +\infty$, Proposition 7.1 gives a control over C_u defined by (5.5):

$$\begin{aligned} C_u^2 &= d^{-1} \exp\left(\frac{u^2 T_d^2}{2}(1 + o(1))\right) (1 - \exp\left(\frac{-u^2 T_d^2}{2}(1 + o(1))\right))^{-1} \\ &= d^{-1 + \frac{u^2}{2}(1 + o(1))}. \end{aligned}$$

Since $u \leq \sqrt{2}$, the exponent of d in C_u is $o(1)$.

Case 1: for the nonzero ξ_j 's, assume that $\limsup(uT_d - \mathbb{E}_{\theta_j}(t_j)) < +\infty$. In this case, the probability $\mathbb{P}_{\theta_j}(t_j > uT_d) = \mathbb{P}_{\theta_j}(t_j - \mathbb{E}_{\theta_j}(t_j) > uT_d - \mathbb{E}_{\theta_j}(t_j))$ is bounded away from zero. This follows from the asymptotic normality of $t_j - \mathbb{E}_{\theta_j}(t_j)$ for $\mathbb{E}_{\theta_j}(t_j) = O(T_d)$ (see Remark 7.1)

Case 2: for the nonzero ξ_j 's, assume that $uT_d - \mathbb{E}_{\theta_j}(t_j) \rightarrow +\infty$. Then, for any nonzero ξ_j , Proposition 7.1 implies that

$$\log \mathbb{P}_{\theta_j}(t_j > uT_d) \geq -\frac{(uT_d - cT_d)^2}{2}(1 + o(1)).$$

Recall that the number of nonzero ξ_j is equal to $K = d^{1-b}$ and that for all nonzero ξ_j , $\mathbb{E}_{\theta_j}(t_j) \geq cT_d$ for some positive c such that $\max_{j: \xi_j=1} \mathbb{E}_{\theta_j}(t_j) = O(T_d)$. To sum up, the cases 1 and 2 entail that

$$\begin{aligned} \mathbb{E}_{\bar{\theta}}(L(u)) &= C_u \sum_{j: \xi_j=1} \left(\mathbb{P}_{\theta_j}(t_j > uT_d) - \tilde{\Phi}_0(uT_d) \right) \\ &\geq C_u K \left(d^{-\frac{((u-c)+)^2}{2}(1+o(1))} - d^{-\frac{u^2}{2}(1+o(1))} \right) \\ &= d^{-\frac{1}{2} + \frac{u^2}{4}(1+o(1)) + 1-b} \left(d^{-\frac{((u-c)+)^2}{2}(1+o(1))} - d^{-\frac{u^2}{2}(1+o(1))} \right) (1 + o(1)) \\ &= d^{\frac{1}{2}-b + (\frac{u^2}{4} - \frac{((u-c)+)^2}{2})(1+o(1))} (1 + o(1)). \end{aligned}$$

Similarly, let us study the variance of $L(u)$. Using Proposition 7.1, we obtain

$$\begin{aligned} \text{Var}_{\bar{\theta}}(L(u)) &= C_u^2 \sum_{j: \xi_j=1} \mathbb{P}_{\theta_j}(t_j > uT_d) \mathbb{P}_{\theta_j}(t_j \leq uT_d) + C_u^2 \sum_{j: \xi_j=0} \tilde{\Phi}_0(uT_d)(1 - \tilde{\Phi}_0(uT_d)) \\ &= C_u^2 K \mathbb{P}_{\theta_j}(t_j > uT_d)(1 + o(1)) + (d^{b-1} + d^{-b})(1 + o(1)) \\ &= (C_u \mathbb{E}_{\bar{\theta}}(L(u)) + d^{b-1})(1 + o(1)) \\ &= o(d^\eta \mathbb{E}_{\bar{\theta}}(L(u))) + o(1), \quad \eta = o(1). \end{aligned}$$

7.2. Upper bound

Remark 7.2. Note that the condition $T_d \max_k w_k(r_\epsilon^*) = o(1)$ follows from assumption $\log d = o(\epsilon^{-2/(2\tau+1)})$. Indeed, Remark 4.2 and relations (4.10) imply that $T_d \max_k w_k(r_\epsilon^*) \leq (r_\epsilon^*)^{1/(2\tau)} T_d$, where the term on the right-hand side goes to zero as soon as $\log d = o(\epsilon^{-2/(2\tau+1)})$. Therefore, assumption $\log d = o(\epsilon^{-2/(2\tau+1)})$ allows us to apply Proposition 7.1 and Lemma 7.1.

Proof of (ii)–Theorem 5.1.

Type I error probability of $\psi_\alpha^{\chi^2}$. It follows from the Central Limit Theorem that, under the null hypothesis, t_b is asymptotically a standard normal random variable. Therefore

$$\mathbb{P}_0(t_b > T_\alpha) = \Phi(-T_\alpha) + o(1) = \alpha + o(1).$$

Type II error probability of $\psi_\alpha^{\chi^2}$ uniformly over $\Theta_d(\tau, r_\epsilon, b)$ for $r_\epsilon \geq Br_\epsilon^*$, $B \geq 1$. Thanks to Lemma 7.1, uniformly over $\bar{\theta} \in \Theta_d(\tau, r_\epsilon, b)$, we have

$$\begin{aligned} \text{Var}_{\bar{\theta}}(t_b) &= \frac{1}{d} \sum_{j=1}^d (1 + O(\mathbb{E}_{\theta_j}(t_{j,b}))) \\ \mathbb{E}_{\bar{\theta}}(t_b) &= d^{-1/2} \sum_{j=1}^d \mathbb{E}_{\theta_j}(t_{j,b}). \end{aligned}$$

This implies that $\text{Var}_{\bar{\theta}}(t_b) = o((\mathbb{E}_{\bar{\theta}}(t_b))^2)$ provided that $\mathbb{E}_{\bar{\theta}}(t_b) \rightarrow +\infty$. Let us study $\mathbb{E}_{\bar{\theta}}(t_b)$: from Proposition 4.1, Lemma 7.1, and relation (4.7), we get uniformly over $\Theta_d(\tau, r_\epsilon, b)$ with $r_\epsilon \geq Br_\epsilon^*$, $B \geq 1$:

$$\mathbb{E}_{\bar{\theta}}(t_b) \geq d^{1/2-b} B^2 a(r_\epsilon^*) \rightarrow +\infty \text{ as soon as } B^2 d^{1/2-b} a(r_\epsilon^*) \asymp B^2 \rightarrow +\infty, \text{ i.e., as soon as } r_\epsilon/r_\epsilon^* \rightarrow +\infty, \quad (7.9)$$

where $r_\epsilon^* \asymp (\epsilon^4 d^{2b-1})^{\tau/(4\tau+1)}$.

Due to (7.9), using Markov's inequality and Lemma 7.1, for all $\bar{\theta}$ in $\Theta_d(\tau, r_\epsilon, b)$,

$$\begin{aligned} \mathbb{P}_{\bar{\theta}}(t_b \leq T_\alpha) &= \mathbb{P}_{\bar{\theta}}(t_b - \mathbb{E}_{\bar{\theta}}(t_b) \leq T_\alpha - \mathbb{E}_{\bar{\theta}}(t_b)) \\ &\leq \mathbb{P}_{\bar{\theta}}(|t_b - \mathbb{E}_{\bar{\theta}}(t_b)| \geq \mathbb{E}_{\bar{\theta}}(t_b) - T_\alpha) \\ &\leq \frac{\text{Var}_{\bar{\theta}}(t_b)}{(\mathbb{E}_{\bar{\theta}}(t_b) - T_\alpha)^2} = o(1). \end{aligned}$$

This entails that $\beta(\psi_\alpha^{\chi^2}, \Theta_d(\tau, r_\epsilon, b))$ goes to zero as soon as $d^{1/2-b} a(r_\epsilon) \rightarrow +\infty$, i.e., as soon as $\frac{a(r_\epsilon)}{a(r_\epsilon^*)} \rightarrow +\infty$ where $a(r_\epsilon^*) \asymp d^{b-1/2}$.

Proof of (ii)–Theorem 5.2.

Type I error probability of ψ^{HC} . Observe that $w(\psi^{HC}) \leq w(\psi^L) + w(\psi^{max})$. The assumption $\log(d) = o(\epsilon^{-2/(2\tau+1)})$ implies that $T_d \max_k w_k(r_\epsilon^*) = o(1)$. Therefore the application of Proposition 7.1 and the fact that $D^2 > 2$ and $N = o(T_d)$ yield

$$\begin{aligned} w(\psi^{max}) &= \mathbb{P}_0(\max_{1 \leq j \leq d} \max_{1 \leq l \leq N} t_{j,b_l} > DT_d) \leq \sum_{j=1}^d \sum_{l=1}^N \mathbb{P}_0(t_{j,b_l} > DT_d) \\ &\leq Nd \exp(-D^2 T_d^2 / 2(1 + o_d(1))) = Nd^{1-D^2/2(1+o_d(1))} \rightarrow 0. \end{aligned}$$

By Lemma 7.1 and applying Markov's inequality,

$$\begin{aligned} w(\psi^L) &= \mathbb{P}_0(\max_{1 \leq l \leq N-1} L(u_l, b_l) > H) \leq \sum_{l=1}^{N-1} \mathbb{P}_0(L(u_l, b_l) > H) \\ &\leq \sum_{l=1}^{N-1} \frac{\text{Var}_0(L(u_l, b_l))}{H^2} \\ &\leq \frac{(N-1)}{H^2}, \end{aligned}$$

which goes to zero as $d \rightarrow +\infty$ since $H \sim (\log d)^C$, with $C > \frac{1}{4}$ and $N = o_d(T_d)$.

Type II error probability of ψ^{HC} uniformly over $\Theta_d(\tau, r_\epsilon, b)$. For any $\bar{\theta} \in \Theta_d(\tau, r_\epsilon, b)$, we obtain

$$\mathbb{E}_{\bar{\theta}}(1 - \psi^{HC}) \leq \min(\mathbb{E}_{\bar{\theta}}(1 - \psi^{max}), \mathbb{E}_{\bar{\theta}}(1 - \psi^L)), \quad (7.10)$$

$$\mathbb{E}_{\bar{\theta}}(1 - \psi^{max}) \leq \min_{j: \xi_j=1} \min_{1 \leq l \leq N} \mathbb{P}_{\theta_j}(t_{j,b_l} \leq DT_d). \quad (7.11)$$

First, let us consider the alternatives $\bar{\theta} \in \Theta_d(\tau, r_\epsilon, b)$ such that for a nonzero ξ_j , there exists $l \in \{1, \dots, N\}$ for which $\mathbb{E}_{\theta_j} t_{j,b_l} \geq D_1 T_d$ with $D_1 > D$. From Lemma 7.1(i) and Markov's inequality, we obtain

$$\begin{aligned} \mathbb{P}_{\theta_j}(t_{j,b_l} \leq DT_d) &\leq \mathbb{P}_{\theta_j}(|t_{j,b_l} - \mathbb{E}_{\theta_j}(t_{j,b_l})| \geq \mathbb{E}_{\theta_j}(t_{j,b_l}) - DT_d) \\ &\leq \frac{\text{Var}_{\theta_j}(t_{j,b_l})}{(\mathbb{E}_{\theta_j}(t_{j,b_l}) - DT_d)^2} = o(1). \end{aligned} \quad (7.12)$$

Second, in view of (7.10), (7.11), (7.12), it suffices to study the test procedures ψ^L under the alternatives $\bar{\theta} \in \Theta_d(\tau, r_\epsilon, b)$ such that $\max_{j: \xi_j=1} \max_{1 \leq l \leq N} \mathbb{E}_{\theta_j} t_{j,b_l} = O(T_d)$. Then we obtain

$$\mathbb{E}_{\bar{\theta}}(1 - \psi^L) = \mathbb{P}_{\bar{\theta}}(\max_{1 \leq l \leq N-1} L(u_l, b_l) \leq H) \leq \min_{1 \leq l \leq N-1} \mathbb{P}_{\bar{\theta}}(L(u_l, b_l) \leq H).$$

For any $l \in \{1, \dots, N-1\}$,

$$\begin{aligned} \mathbb{P}_{\bar{\theta}}(L(u_l, b_l) \leq H) &\leq \mathbb{P}_{\bar{\theta}}(L(u_l, b_l) - \mathbb{E}_{\bar{\theta}}(L(u_l, b_l)) \leq H - \mathbb{E}_{\bar{\theta}}(L(u_l, b_l))) \\ &\leq \mathbb{P}_{\bar{\theta}}(-|L(u_l, b_l) - \mathbb{E}_{\bar{\theta}}(L(u_l, b_l))| \leq H - \mathbb{E}_{\bar{\theta}}(L(u_l, b_l))) \\ &\leq \mathbb{P}_{\bar{\theta}}(|L(u_l, b_l) - \mathbb{E}_{\bar{\theta}}(L(u_l, b_l))| \geq -H + \mathbb{E}_{\bar{\theta}}(L(u_l, b_l))) \\ &\leq \frac{\text{Var}_{\bar{\theta}}(L(u_l, b_l))}{(\mathbb{E}_{\bar{\theta}}(L(u_l, b_l)) - H)^2}. \end{aligned} \quad (7.13)$$

For any $b_l \in (1/2, 1)$, if we prove that $\inf_{\bar{\theta} \in \Theta_d(\tau, r_\epsilon, b)} \mathbb{E}_{\bar{\theta}}(L(u_l, b_l))$ goes to infinity as a power of d ($d \rightarrow +\infty$), then Lemma 7.1 and the choice of H (recall $H = O_d((\log d)^C)$, with $C > 1/4$) yield the result since in this case the right-hand side of relation (7.13) goes to zero.

Third, for $b \in (1/2, 1)$, take an index l in $\{1, \dots, N-1\}$ such that $b_l \leq b \leq b_{l+1}$. This, combined with the continuity of φ , yields

$$b_l = b + o(1), \quad r_\epsilon^*(b_l) \leq r_\epsilon^*(b) \sim r_\epsilon^*(b_l), \quad a(r_\epsilon^*(b_l)) \leq a(r_\epsilon^*(b)) \sim a(r_\epsilon^*(b_l)).$$

Let $\bar{\theta} \in \Theta_d(\tau, r_\epsilon, b)$ with $b \in (1/2, 1)$ and $\liminf(a(r_\epsilon)/a(r_\epsilon^*(b))) > 1$. Then $r_\epsilon \geq (1 + \delta)r_\epsilon^*(b_l)$ for some $\delta > 0$. Proposition 4.1 entails that for j such that $\xi_j = 1$ we have

$$\mathbb{E}_{\theta_j} t_{j,b_l} \geq (1 + \delta)^2 a(r_\epsilon^*(b_l)) \sim (1 + \delta)^2 a(r_\epsilon^*(b)) \sim (1 + \delta)^2 \varphi(b) T_d.$$

We then derive from Lemma 7.1 with $c = c(b) = (1 + \delta)^2 \varphi(b)$ that

$$\inf_{\bar{\theta} \in \Theta_d(\tau, r_\epsilon, b)} \mathbb{E}_{\bar{\theta}}(L(u_l, b_l)) > d^{\frac{1}{2} + \frac{u^2}{4} - b - \frac{((u_l - c(b))_+)^2}{2}} (1 + o(1)) (1 + o(1)). \quad (7.14)$$

Finally, denote the main term in the exponent of d in (7.14) by

$$M = \frac{1}{2} + \frac{u(b)^2}{4} - b - \frac{((u(b) - c(b))_+)^2}{2}.$$

To obtain the result, it is sufficient to prove that M is positive and bounded away from zero for any $\delta > 0$.

Intermediate sparsity case. This case corresponds to $b \in (1/2, 3/4]$. Recall that $u(b) = 2\varphi_1(b)$, where φ_1 is defined in (2.2). Then

$$M > 0 \Leftrightarrow \begin{cases} (\varphi_1^2(b)/2)((1+\delta)^2 - 1)(3 - (1+\delta)^2) > 0 & \text{for } 0 < \delta < \sqrt{2} - 1 \\ \varphi_1^2(b)/2 > 0 & \text{for } \delta \geq \sqrt{2} - 1 \end{cases}.$$

The latter inequalities are obviously satisfied. This leads to the result.

Highest sparsity case. In this case $b \in (3/4, 1)$ and $u(b) = \sqrt{2}$. Then

$$M > 0 \Leftrightarrow \begin{cases} ((1+\delta)^2 - 1)\varphi_2(b) > 0 & \text{for } (1+\delta)^2\varphi_2(b) \leq \sqrt{2} \\ 1 - b > 0 & \text{for } (1+\delta)^2\varphi_2(b) > \sqrt{2} \end{cases}.$$

Again, the latter inequalities are satisfied, and the result follows.

Proof of (i)–Theorem 6.1.

Similar to the proof of part (ii) of Theorem 5.1, due to (6.3) and (6.4), uniformly over $\bar{\theta} \in \Theta_d^{ext}(\tau, K^{1/2}, K^{1/2}r_\epsilon, b)$, the type II error probability of $\psi_\alpha^{\chi^2}$ goes to zero as soon as $r_\epsilon/r_\epsilon^* \rightarrow +\infty$.

Proof of (ii)–Theorem 6.1.

The proof of the fact that the type II error probability of ψ^{HC} goes to zero as $d \rightarrow +\infty$ is similar to the one of Theorem 5.2. Recall that $K = d^{1-b}$ is the number of nonzero ξ_j 's and suppose without loss of generality that $\xi_j = 1, \forall j \in \{1, \dots, K\}$ and $\xi_j = 0, \forall j \in \{K+1, \dots, d\}$. Note that relations (7.10) and (7.11) remain valid for any $\bar{\theta} \in \Theta_d^{ext}(\tau, K^{1/2}, K^{1/2}r_\epsilon, b)$.

First, similarly to (7.12), for any $\bar{\theta} \in \Theta_d^{ext}(\tau, K^{1/2}, K^{1/2}r_\epsilon, b)$ such that for the nonzero ξ_j 's, there exists $l \in \{1, \dots, N\}$ for which $\mathbb{E}_{\theta_j} t_{j,b_l} \geq D_1 T_d$ with $D_1 > D$, the type II error probability of ψ^{HC} vanishes asymptotically. Therefore, it suffices to study the test procedures ψ^L under the alternatives $\bar{\theta} \in \Theta_d^{ext}(\tau, K^{1/2}, K^{1/2}r_\epsilon, b)$ such that $\max_{j: \xi_j=1} \max_{1 \leq l \leq N} \mathbb{E}_{\theta_j} t_{j,b_l} = O(T_d)$. Therefore, let us take $\delta > 0$ and consider the alternatives that are as far away from the null hypothesis as r_ϵ such that $r_\epsilon \geq (1+\delta)r_\epsilon^*(b)$, where $r_\epsilon^*(b)$ is determined by $a(r_\epsilon^*(b)) \sim T_d \varphi(b)$.

Second, for any $l \in \{1, \dots, N\}$, observe that the only difference between the proofs of the extended and initial problems lies in the study of

$$\inf_{\bar{\theta} \in \Theta_d^{ext}(\tau, K^{1/2}, K^{1/2}r_\epsilon, b)} \sum_{j=1}^K \mathbb{P}_{\theta_j}(t_{j,b_l} - \mathbb{E}_{\theta_j}(t_{j,b_l}) > u_l T_d - \mathbb{E}_{\theta_j}(t_{j,b_l})). \quad (7.15)$$

Now it is no more possible to control (7.15) by using Lemma 7.1 (ii) because the condition $\mathbb{E}_{\theta_j}(t_j) \geq cT_d$ is not necessarily satisfied for all nonzero ξ_j 's. In fact, the only condition we have is

$$\sum_{j=1}^K \mathbb{E}_{\theta_j}(t_j) \geq cKT_d \text{ with some constant } c > 1.$$

Let us now explain why the current proof is reduced to the study of (7.15). As in (7.13), we get for any $\bar{\theta}$ in $\Theta_d^{ext}(\tau, K^{1/2}, K^{1/2}r_\epsilon, b)$,

$$\mathbb{P}_{\bar{\theta}}(\max_{1 \leq l \leq N} L(u_l, b_l) \leq H) \leq \min_{1 \leq l \leq N} \frac{\text{Var}_{\bar{\theta}}(L(u_l, b_l))}{(\mathbb{E}_{\bar{\theta}}(L(u_l, b_l)) - H)^2}.$$

Due to Lemma 7.1 and the fact that $H = O_d((\log d)^C)$ with $C > 1/4$, in order to obtain the result, it remains to prove that for any l such that $b_l \leq b \leq b_{l+1}$, $\inf_{\bar{\theta} \in \Theta_d^{ext}(\tau, K^{1/2}, K^{1/2}r_\epsilon, b)} \mathbb{E}_{\bar{\theta}}(L(u_l, b_l)) \xrightarrow{d \rightarrow +\infty} +\infty$ as a positive power of d . Finally, recall that

$$\mathbb{E}_{\bar{\theta}}(L(u_l, b_l)) = C_{u_l, b_l} \sum_{j=1}^K \left(\mathbb{P}_{\theta_j}(t_{j,b_l} - \mathbb{E}_{\theta_j}(t_{j,b_l}) > u_l T_d - \mathbb{E}_{\theta_j}(t_{j,b_l})) - \tilde{\Phi}_{0, b_l}(u_l T_d) \right) \quad (7.16)$$

where $C_{u_l, b_l} = (d\tilde{\Phi}_{0, b_l}(u_l T_d)(1 - \tilde{\Phi}_{0, b_l}(u_l T_d)))^{1/2}$ and $\tilde{\Phi}_{0, b_l}(x) = \mathbb{P}_0(t_{j, b_l} > x)$. The term on the right-hand side of (7.16) corresponds to the product of (7.15) and C_{u_l, b_l} . The quantity C_{u_l, b_l} is controlled by Lemma 7.1 and Proposition 7.1. Thus it remains to study (7.15).

Third, the application of Proposition 7.1 gives the following approximation of (7.15),

$$\sum_{j=1}^K \mathbb{P}_{\theta_j}(t_{j,b_l} - \mathbb{E}_{\theta_j}(t_{j,b_l}) > u_l T_d - \mathbb{E}_{\theta_j}(t_{j,b_l})) = \sum_{j=1}^K \exp\left(-\frac{((u_l T_d - \mathbb{E}_{\theta_j}(t_{j,b_l}))_+)^2}{2}\right) O(1).$$

Recall that $a(r_\epsilon)$ given by (4.8) is the solution of the extremal problem (4.1). Set $\eta_j = \mathbb{E}_{\theta_j}(t_{j,b_l})$, $\eta_0 = (1+\delta)^2 a(r_\epsilon^*(b)) \sim (1+\delta)^2 a(r_\epsilon^*(b_l))$ and $f_T(\eta) = \exp(-\frac{(T-\eta)^2}{2}) \forall \eta \in [0, R]$, where $R = R(T) > 0$ will be specified later on. Consider also

$$F_{K,T}(\eta_0) \triangleq \inf \sum_{j=1}^K f_T(\eta_j) \text{ subject to } \sum_{j=1}^K \eta_j \geq K\eta_0.$$

Due to relation (6.4), we have for the sequence $w(r_\epsilon^*(b_l))$ that

$$\sum_{j=1}^K \eta_j = \sum_{j=1}^K \mathbb{E}_{\theta_j}(t_{j,b_l}) = \frac{1}{\epsilon^2} \sum_{j=1}^K \kappa(\theta_j, w_l) \geq K(1+\delta)^2 a(r_\epsilon^*(b_l)) \sim K\eta_0.$$

Then, in order to obtain the same right-hand side as in (7.14), it is sufficient to show that for any l in $\{1, \dots, N\}$ such that $T = u_l T_d$, relation (7.17) which is stated below, holds:

$$F_{K,T}(\eta_0) = K f_T(\eta_0). \quad (7.17)$$

This is handled by a technical result similar to the one stated in Lemma 7.4 and Lemma 7.5 in Ingster *et al.* [17]. The proof of Lemma 7.2 is postponed to Section 7.4.

Lemma 7.2. Set $\lambda = (T - \eta_0) f_T(\eta_0)$.

$$\text{If } 0 < \eta_0 < T - 1 \text{ and } T < R < T + ((T - \eta_0)^2 - 2 \log(1 + 2(T - \eta_0)^2))^{1/2}, \quad (7.18)$$

then

$$\inf_{\eta \in [0, R]} (f_T(\eta) - \lambda \eta) = f_T(\eta_0) - \lambda \eta_0, \quad (7.19)$$

which implies that

$$F_{K,T}(\eta_0) = K f_T(\eta_0). \quad (7.20)$$

As $d \rightarrow +\infty$, for any $l \in \{1, \dots, N\}$ such that $T = u_l T_d$ with $u_l > (1+\delta)^2 \varphi(b)$ and $R = p T_d$ with $u_l < p < u_l + \frac{u_l - (1+\delta)^2 \varphi(b)}{2}$, the conditions in (7.18) are then satisfied. Therefore the application of Lemma 7.2 yields the results since for all $\bar{\theta} \in \Theta_d^{ext}(\tau, K^{1/2}, K^{1/2} r_\epsilon, b)$,

$$\mathbb{E}_{\bar{\theta}}(L(u_l, b_l)) > C_{u_l, b_l} K \left(\exp\left(-\frac{((u_l T_d - (1+\delta)^2 a(r_\epsilon^*(b_l)))_+)^2}{2}\right) O(1) - \exp\left(-\frac{u_l T_d}{2}(1 + o(1))\right) \right),$$

which corresponds to the right-hand side of (7.14).

7.3. Lower Bound

The prior distribution we consider is a classical one for a functional Gaussian model. In Section 4.3 of [13] it is referred to as the symmetric Three-point Factors.

PRIOR. Before defining the prior Π^d formally, we shall start with an informal discussion.

The prior Π^d adds mass on $(\xi_j \theta_j)_{1 \leq j \leq d}$: the components are i.i.d. and ξ_j and θ_j are supposed to be independent. A natural choice for ξ_j is a Bernoulli with a parameter $p_d \in (0, 1)$ such that $\mathbb{E}(\sum_{j=1}^d \xi_j) \sim K$. The θ_j 's are binary random variables (with probability 1/2) such that $\theta_j^2 = (\theta^*)^2$ where the sequence θ^* is a solution of the extremal problem (3.1); this guarantees that θ_j belongs to $\Theta(\tau, r_\epsilon)$.

Now, we define the prior distribution more precisely. Let ρ_d be any sequence of positive numbers such that $\rho_d \xrightarrow{d \rightarrow +\infty} 0$ and $d^{1-b} \rho_d^s \xrightarrow{d \rightarrow +\infty} +\infty$, $\forall b \in (0, 1)$, $\forall s > 0$. Consider two sequences $(\xi_j)_j$ and $(\theta_{j,k})_{j,k}$ of independent random variables whose distributions are the following:

$$\begin{cases} \xi_j \sim \text{Bernoulli } B(p_d) \text{ with } p_d = d^{-b}(1 + \rho_d), j \in \{1, \dots, d\}, \\ \theta_{j,k} = \varepsilon_{j,k} \in z_k, \text{ with } \mathbb{P}(\varepsilon_{j,k} = 1) = \mathbb{P}(\varepsilon_{j,k} = -1) = \frac{1}{2}, j \in \{1, \dots, d\}, k \in \mathbf{Z}. \end{cases}$$

The sequence $(z_k)_{k \in \mathbf{Z}}$ is deterministic and is defined as follows: $(\varepsilon z_k)_k = (\theta_k^*)_{k \in \mathbf{Z}} = \theta^*$ where θ^* is the sequence that leads to the solution (4.8) of the extremal problem (4.1). In particular, this entails that

$$\sum_{k \in \mathbf{Z}} \frac{z_k^4}{2} = a^2(r_\epsilon), \quad (7.21)$$

$$(2\pi)^{2\tau} \sum_{k \in \mathbf{Z}} |k|^{2\tau} (\varepsilon z_k)^2 \leq 1, \quad (7.22)$$

$$\sum_{k \in \mathbf{Z}} (\varepsilon z_k)^2 \geq r_\epsilon^2. \quad (7.23)$$

The sequences $(\xi_j)_j$ and $(\theta_{j,k})_{j,k}$ are also taken mutually independent. For each j in $\{1, \dots, d\}$, we define the prior distribution π_j^d on (ξ_j, θ_j) as follows:

$$\pi_j^d = (1 - p_d)\delta_0 + p_d \prod_{k \in \mathbf{Z}} \pi_{j,k} = (1 - p_d)\delta_0 + p_d \pi_j, \quad (7.24)$$

where $\pi_j = \prod_{k \in \mathbf{Z}} \pi_{j,k}$, $\pi_{j,k} = \frac{1}{2}(\delta_{(-\varepsilon z_k)} + \delta_{(\varepsilon z_k)})$ puts mass on $\theta_{j,k}$ and δ is the Dirac mass. Finally, we define the global prior Π^d by

$$\Pi^d = \prod_{j=1}^d \pi_j^d.$$

MINIMAX AND BAYESIAN RISKS. Denote by \mathbb{P}_{Π^d} the mixture of the measures \mathbb{P}_{θ} over the prior Π^d , and let $\gamma(Q)$ be the minimal total error probability for testing a simple null hypothesis $H_0 : \mathbb{P} = \mathbb{P}_0$ against a simple alternative $H_1 : \mathbb{P} = Q$ regarding the measure \mathbb{P} of our observations $(x_{j,k})_{k \in \mathbf{Z}, 1 \leq j \leq d}$.

Proposition 7.2.

$$\gamma \geq \gamma(\mathbb{P}_{\Pi^d}) + o(1), \quad (7.25)$$

where γ is the minimax total error probability over $\Theta_d(\tau, r_\epsilon, b)$ (see (3.2)).

Proof of Proposition 7.2.

Consider two sets $\Xi(s)$ and $\Xi^+(s)$ defined by

$$\Xi(s) = \{\zeta = \varepsilon (\xi_1(\varepsilon_{1,k} z_k)_k, \dots, \xi_d(\varepsilon_{d,k} z_k)_k) : \sum_{j=1}^d \xi_j = s\}, \quad 0 \leq s \leq d, \quad \Xi^+(s) = \bigcup_{s \leq l \leq d} \Xi(l).$$

First, due to relations (7.22) and (7.23), $\Xi(K)$ is included in $\Theta_d(\tau, r_\epsilon, b)$. This entails that

$$\gamma \geq \gamma(\Xi(K)). \quad (7.26)$$

Second, let us introduce some additional priors: for any subset $u \subset \{1, \dots, d\}$, define $\pi_u = \prod_{j \in u} \pi_j \prod_{j \notin u} \delta_0$, where π_j is as in (7.24). Note that π_u has a support on the collections $\zeta = \varepsilon (\xi_1(\varepsilon_{1,k} z_k)_k, \dots, \xi_d(\varepsilon_{d,k} z_k)_k)$ with $\xi_j = 1$ if and only if $j \in u$. For any integer s such that $0 \leq s \leq d$, let $\mathcal{G}_{d,s}$ be the set of all subsets $u \subset \{1, \dots, d\}$ of cardinality s , and define $\pi_{(s)}$ as the uniform distribution on $\mathcal{G}_{d,s}$:

$$\pi_{(s)} = \frac{1}{\binom{d}{s}} \sum_{u \in \mathcal{G}_{d,s}} \pi_u.$$

Observe that the prior Π^d is of the form $\Pi^d = \sum_{s=0}^d r_s \pi(s)$ where $r_s = p_d^s (1 - p_d)^{d-s}$. Clearly, $\pi_{(K)}(\Xi(K)) = 1$, which implies

$$\gamma(\Xi(K)) \geq \gamma(\mathbb{P}_{\pi_{(K)}}). \quad (7.27)$$

Third, consider the conditional prior of the form Π_+^d with respect to $\Xi(K)^+$, i.e., $\Pi_+^d(A) = \frac{\Pi^d(A \cap \Xi(K)^+)}{\Pi^d(\Xi(K)^+)}$ which is of the form $\Pi_+^d = \sum_{s=K}^d q_s \pi(s)$ with $q_s = \frac{r_s}{\sum_{s=K}^d r_s}$, $K \leq s \leq d$.

Let us prove that

$$\gamma(\mathbb{P}_{\pi_{(K)}}) \geq \gamma(\mathbb{P}_{\Pi_+^d}). \quad (7.28)$$

Denote by $\mathcal{X}_K = \{(x_{j,k})_{j,k} : \frac{d\mathbb{P}_{\pi_{(K)}}}{d\mathbb{P}_0}((x_{j,k})_{j,k}) < 1\}$ the admissible set of the optimal test for testing $H_0 : \mathbb{P} = \mathbb{P}_0$ against $H_1 : \mathbb{P} = \mathbb{P}_{\pi_{(K)}}$. Since

$$\gamma(\mathbb{P}_{\pi_{(K)}}) = 1 - \mathbb{P}_0(\mathcal{X}_K) + \mathbb{P}_{\pi_{(K)}}(\mathcal{X}_K) \quad \text{and} \quad \gamma(\mathbb{P}_{\Pi_+^d}) \leq 1 - \mathbb{P}_0(\mathcal{X}_K) + \mathbb{P}_{\Pi_+^d}(\mathcal{X}_K),$$

proving (7.28) is then reduced to checking that

$$\mathbb{P}_{\pi_{(K)}}(\mathcal{X}_K) \geq \mathbb{P}_{\Pi_+^d}(\mathcal{X}_K). \quad (7.29)$$

In view of Proposition 2.5 in [13], \mathcal{X}_K is a convex set. Also, the set \mathcal{X}_K is sign-invariant and invariant with respect to all permutations of the $x_{j,k}$'s; the measures $\mathbb{P}_{\pi(s)}$, $0 \leq s \leq d$ have the same property of invariance with respect to all permutations of the $x_{j,k}$'s. These observations imply

$$\mathbb{P}_{\pi_{(K)}}(\mathcal{X}_K) = \mathbb{P}_{\bar{\theta}^K}(\mathcal{X}_K), \quad \mathbb{P}_{\Pi_+^d}(\mathcal{X}_K) = \sum_{s=K}^d q_s \mathbb{P}_{\bar{\theta}^s}(\mathcal{X}_K),$$

where $\bar{\theta}^s = \epsilon(\underbrace{z, \dots, z}_s, 0, \dots, 0)$, $z = (z_k)_{k \in \mathbb{Z}}$. Since $\theta_{j,k}^s \geq \theta_{j,k}^K \geq 0$, $\forall j, k, s \geq K$, the application of Lemma 2.4 in [13] entails that $\mathbb{P}_{\bar{\theta}^K}(\mathcal{X}_K) \geq \mathbb{P}_{\bar{\theta}^s}(\mathcal{X}_K)$, $s \geq K$. This yields relation (7.29) and hence relation (7.28).

Finally, in view of Proposition 2.11 in [13], it remains to check that

$$\gamma(\mathbb{P}_{\Pi_+^d}) = \gamma(\mathbb{P}_{\Pi^d}) + o(1). \quad (7.30)$$

Similarly to the proof of Proposition 2.9. in [13], it is easily seen that (7.30) follows from the relation

$$\Pi^d(\Xi^+(K)) \xrightarrow{d \rightarrow +\infty} 1. \quad (7.31)$$

Acting as in the proof of Proposition 3 in [12], we obtain by Chebyshev's inequality,

$$\begin{aligned} 1 - \Pi^d(\Xi^+(K)) &= \Pi^d(\sum \xi_j < d^{1-b}) \\ &= \Pi^d(dp_d - \sum \xi_j > dp_d - d^{1-b}) \\ &\leq \frac{d^{1-b}(1 + \rho_d)(1 - d^{-b}(1 + \rho_d))}{(d^{1-b}\rho_d)^2}, \end{aligned}$$

where the ratio on the right-hand side tends to zero as d goes to infinity. Relation (7.31) is then proved.

As relations (7.26), (7.27), (7.28), and (7.30) imply (7.25), the proof of Proposition 7.2 is completed.

Due to Proposition 7.2. the proof of the lower bound is reduced to bounding $\gamma^* \triangleq \gamma(\mathbb{P}_{\Pi^d})$ from below.

Before studying γ^* , we introduce some useful notation and make some helpful remarks. Denote by $\|\cdot\|_{TV}$ and $\|\cdot\|_2$ the distance in variation and the L_2 -distance between any pair of probabilities (P, Q) ; the latter one is defined by

$$\|P - Q\|_2^2 = \begin{cases} +\infty & \text{if } P \text{ does not dominate } Q, \\ \mathbb{E}_P(L - 1)^2 & \text{if } P \text{ dominates } Q, \end{cases} \quad (7.32)$$

where $L = \frac{dQ}{dP}$ is the Radon-Nikodym derivative of Q with respect to P .

Remark 7.3. *Note that*

- $\|\cdot\|_{TV} = \|\cdot\|_1$, where $\|\cdot\|_1$ is the L_1 -distance.
- If P dominates Q , then $\|P - Q\|_2^2 = \mathbb{E}_P(L^2) - 1$.
- As stated in Proposition 2.12 of [13],
 If $\|P - Q\|_2 = o(1)$, then $\|P - Q\|_1 = o(1)$,
 If $\|P - Q\|_2$ is bounded, then $\limsup \|P - Q\|_1 < 2$.

Using Remark 7.3, one has

$$\text{If } \|P_0 - P_{\Pi^d}\|_2 = o(1) \text{ then } \|P_0 - P_{\Pi^d}\|_1 = o(1) \text{ and } \gamma^* \rightarrow 1. \quad (7.33)$$

$$\text{If } \|P_0 - P_{\Pi^d}\|_2 = O(1) \text{ then } \limsup \|P_0 - P_{\Pi^d}\|_1 < 2 \text{ and } \liminf \gamma^* > 0. \quad (7.34)$$

Therefore, if needed, the L_2 -distance can be conveniently used instead of the total variation distance.

Due to (7.33) and (7.34), it remains to study $\|P_0 - P_{\Pi^d}\|_2$ which is expressed in terms of the Bayesian likelihood ratio $L_{\Pi^d} = \frac{dP_{\Pi^d}}{dP_0}$ (see relation (7.32)).

LIKELIHOOD RATIOS. Here and below, when it is not absolutely necessary, we omit the arguments of the likelihood ratios. Then, observe that L_{Π^d} is defined by:

$$\begin{aligned} L_{\Pi^d} &= \int \prod_{j=1}^d \left(\frac{dP_{\theta_j}}{dP_0} \right) d\Pi^d \\ &= \prod_{j=1}^d \int \left(\frac{dP_{\theta_j}}{dP_0} \right) d\pi_j^d \\ &= \prod_{j=1}^d (1 - p_d + p_d L_j), \end{aligned}$$

where L_j is the likelihood ratio between P_{π_j} and P_0 . Denote also by $L_{\pi_j^d}$ the likelihood ratio between $P_{\pi_j^d}$ and P_0 , i.e., $L_{\pi_j^d} = (1 - p_d + p_d L_j)$. Then L_j is such that

$$\begin{aligned} L_j(x_j) &= \int \prod_{k \in \mathbf{Z}} \left(\frac{dP_{\theta_{j,k}}}{dP_0}(x_{j,k}) \right) d\pi_j \\ &= \prod_{k \in \mathbf{Z}} \frac{1}{2} \left(\exp\left(-\frac{z_k^2}{2} + z_k x_{j,k}/\epsilon\right) + \exp\left(-\frac{z_k^2}{2} - z_k x_{j,k}/\epsilon\right) \right) \\ &= \prod_{k \in \mathbf{Z}} \exp\left(-\frac{z_k^2}{2}\right) \cosh(z_k x_{j,k}/\epsilon), \end{aligned} \quad (7.35)$$

where \cosh is the hyperbolic cosine. Using routine calculations, in particular, using twice the

inequality $1 + x \leq \exp(x)$, $\forall x \in \mathbf{R}$, we obtain

$$\begin{aligned} \mathbb{E}_0(L_{\pi_j^d}^2) &= 1 + p_d^2 \{ \mathbb{E}_0(L_j^2) - 1 \} \\ &= 1 + p_d^2 \left\{ \prod_{k \in \mathbf{Z}} (1 + 2(\sinh(\frac{z_k^2}{2}))^2) - 1 \right\} \\ &\leq 1 + p_d^2 \left\{ \exp\left(\sum_{k \in \mathbf{Z}} 2(\sinh(\frac{z_k^2}{2}))^2\right) - 1 \right\} \\ &\leq \exp(p_d^2 \{ \exp(\sum_{k \in \mathbf{Z}} 2(\sinh(\frac{z_k^2}{2}))^2) - 1 \}), \end{aligned}$$

where \sinh denotes the hyperbolic sine. In view of Remark 7.3, in order to study $\|\mathbb{P}_0 - \mathbb{P}_{\Pi^d}\|_2$, it suffices to study $\mathbb{E}_0(L_{\Pi^d}^2 - 1)^2$. The latter includes the quantity $\mathbb{E}_0(L_{\Pi^d}^2)$ that satisfies

$$\mathbb{E}_0(L_{\Pi^d}^2) = \prod_{j=1}^d \mathbb{E}_0(L_{\pi_j^d}^2) \leq \exp \left(dp_d^2 \{ \exp(\sum_{k \in \mathbf{Z}} 2(\sinh(\frac{z_k^2}{2}))^2) - 1 \} \right). \quad (7.36)$$

As d goes to infinity, the right-hand side of (7.36) goes to one provided that

$$d p_d^2 (\exp(A) - 1) \xrightarrow{d \rightarrow +\infty} 0 \quad \text{with} \quad A = \sum_{k \in \mathbf{Z}} 2(\sinh(\frac{z_k^2}{2}))^2. \quad (7.37)$$

Proof of (i)–Theorem 5.1.

Recall that by assumption $b \in (0, 1/2]$. We shall distinguish between two cases depending on the values of r_ϵ with respect to r_ϵ^* defined in (5.9).

Case 1: $r_\epsilon/r_\epsilon^* = O(1)$. Since $dp_d^2(a(r_\epsilon^*))^2 = O(1)$, it follows that $dp_d^2 a^2(r_\epsilon) = O(1)$. Since dp_d^2 is bounded away from zero, $a^2(r_\epsilon) = O(1)$, and, due to Remark 4.2 and relations (4.10), we have $\sup_k z_k^2 = o(1)$. This entails that $\sinh^2(\frac{z_k^2}{2}) \sim \frac{z_k^4}{4}$, which, due to (7.21), implies that $A \sim \sum \frac{z_k^4}{2} \sim a^2(r_\epsilon)$, and hence $A = O(1)$. It now follows that $\exp(A) - 1 \asymp A$. Finally, we get

$$dp_d^2 (\exp(A) - 1) \asymp dp_d^2 a^2(r_\epsilon) = O(1), \quad (7.38)$$

and the second part of (i) in Theorem 5.1 is proved.

Case 2: $r_\epsilon/r_\epsilon^* = o(1)$. Due to (7.38), we have $dp_d^2 (\exp(A) - 1) \asymp dp_d^2 a^2(r_\epsilon)$, and since $\frac{a^2(r_\epsilon)}{a^2(r_\epsilon^*)} = o(1)$, relation (7.37) is trivially fulfilled.

Proof of (i)–Theorem 5.2.

Now by assumption $b \in (1/2, 1)$. Due to the condition $\log(d) = o(\epsilon^{-2/(2\tau+1)})$, Remark 4.2, and relations (4.10), $\sup_k z_k^2 = o(1)$. As in the moderate case, this yields $A \sim a^2(r_\epsilon)$, and thus we obtain

$$d p_d^2 (\exp(A) - 1) = dp_d^2 \exp(a^2(r_\epsilon)(1 + o(1))). \quad (7.39)$$

Again, we shall consider two cases depending on the values of r_ϵ with respect to r_ϵ^* , where r_ϵ^* is now defined by (5.11).

Case 1: Suppose that $r_\epsilon/r_\epsilon^* = o(1)$. Then $a(r_\epsilon) = o(T_d)$. Due to equation (7.39), this implies that relation (7.37) is fulfilled.

Case 2: Suppose that $r_\epsilon/r_\epsilon^* = O(1)$ and let $c(r_\epsilon)$ be a positive constant satisfying $c^2(r_\epsilon) \log(d) = a^2(r_\epsilon)$. Then the right-hand side of (7.39) can be rewritten as follows:

$$\begin{aligned} dp_d^2 \exp(a^2(r_\epsilon)) &= d^{1-2b} (1 + \rho_d)^2 \exp(\log(d) c^2(r_\epsilon) (1 + o(1))) \\ &= d^{1-2b+c^2(r_\epsilon)(1+o(1))} (1 + \rho_d)^2. \end{aligned}$$

Therefore, relation (7.37) is fulfilled provided that $c(r_\epsilon) < \sqrt{2b-1} = \varphi_1(b)$, where φ_1 is defined in (2.2). This means that a successful detection is impossible if $c(r_\epsilon) < \varphi_1(b)$, which corresponds to the intermediate sparsity case; in fact, the inequality $c(r_\epsilon) < \varphi_1(b)$ is valid for any $b \in (1/2, 1)$ but it could be improved for $b \in (3/4, 1)$. Indeed, for $b \in (3/4, 1)$, one can show that a successful detection is impossible if $c(r_\epsilon)$ is such that $c(r_\epsilon) < \varphi_2(b)$, where the function φ_2 is defined in (2.2), and for $b \in (3/4, 1)$, $\varphi_1(b) < \varphi_2(b)$. That is why the improvement is possible and is achieved by dealing with a truncated version of the Bayesian likelihood ratio L_{Π^d} . From now, let us consider $a(r_\epsilon) = c(r_\epsilon)\sqrt{\log d}$ with $\frac{1}{\sqrt{2}} < c(r_\epsilon) < \sqrt{2}$. The case $c(r_\epsilon) \leq \frac{1}{\sqrt{2}}$ coincides with the intermediate sparsity case when $b \in (1/2, 3/4]$.

Thus, for some positive v , let us define \hat{L}_{Π^d} , the truncated likelihood ratio of L_{Π^d} :

$$\hat{L}_{\Pi^d} = \prod_{j=1}^d \hat{L}_{\pi_j^d} = \prod_{j=1}^d (L_{\pi_j^d}) \mathbb{I}_{(\tilde{l}_j \leq a(r_\epsilon)\sqrt{(2+v)\log d})}, \quad (7.40)$$

where

$$\tilde{l}_j = \log(L_j) + \frac{1}{2} a^2(r_\epsilon). \quad (7.41)$$

Also put

$$l_j = \log(L_j), \quad (7.42)$$

where L_j is defined by (7.35). Now we introduce two new probability measures \mathbb{P}_{ν_j} and \mathbb{P}_{μ_j} expressed in terms of \mathbb{P}_0 as follows:

$$\frac{d\mathbb{P}_{\nu_j}}{d\mathbb{P}_0} = \frac{\exp(l_j)}{\mathbb{E}_0(L_j)}, \quad (7.43)$$

$$\frac{d\mathbb{P}_{\mu_j}}{d\mathbb{P}_0} = \frac{\exp(2l_j)}{\mathbb{E}_0(L_j^2)}. \quad (7.44)$$

In order to get a lower bound for the minimax total error probability, it is sufficient to prove (see the proof of Theorem 4.1 in [11]) that $\mathbb{E}_0((\hat{L}_{\Pi^d} - 1)^2) = o(1)$, where \hat{L}_{Π^d} is defined in (7.40) provided that

$$\mathbb{P}_0\left(\bigcap_{j=1}^d \{\tilde{l}_j \leq a(r_\epsilon)\sqrt{(2+v)\log d}\}\right) \rightarrow 1. \quad (7.45)$$

In fact, it is enough to prove that

$$\sum_{j=1}^d \mathbb{P}_0(\tilde{l}_j > a(r_\epsilon)\sqrt{(2+v)\log d}) \rightarrow 0. \quad (7.46)$$

Relation (7.46), and hence relation (7.45), follows from relation (7.47), which is a part of the next lemma whose proof is postponed to Section 7.4.

Lemma 7.3. *Assume that $r_\epsilon \rightarrow 0$ and $\log d = o(\epsilon^{-2/(2\tau+1)})$. If $T > 0$ is such that $T = O(a^2(r_\epsilon))$, then*

$$\mathbb{P}_0(\tilde{l}_j > T) \leq \exp\left(-\frac{T^2}{2a^2(r_\epsilon)} + o(a^2(r_\epsilon))\right). \quad (7.47)$$

Moreover, if $\liminf(T/a^2(r_\epsilon)) > 1$, then

$$\mathbb{P}_{\nu_j}(\tilde{l}_j > T) \leq \exp\left(-\frac{(T - a^2(r_\epsilon))^2}{2a^2(r_\epsilon)} + o(a^2(r_\epsilon))\right), \quad (7.48)$$

and if $\limsup(T/a^2(r_\epsilon)) < 2$, then

$$\mathbb{P}_{\mu_j}(\tilde{l}_j \leq T) \leq \exp\left(-\frac{(T - 2a^2(r_\epsilon))^2}{2a^2(r_\epsilon)} + o(a^2(r_\epsilon))\right). \quad (7.49)$$

Next, it remains to prove that $\mathbb{E}_0(\hat{L}_{\Pi^d}) \rightarrow 1$ and $\mathbb{E}_0((\hat{L}_{\Pi^d})^2) \rightarrow 1$. This will entail the expected result that $\mathbb{E}_0((\hat{L}_{\Pi^d} - 1)^2) = o(1)$.

First, consider the term $\mathbb{E}_0(\hat{L}_{\Pi^d})$:

$$\begin{aligned} \mathbb{E}_0(\hat{L}_{\Pi^d}) &= \prod_{j=1}^d \mathbb{E}_0(\hat{L}_{\pi_j^d}) \\ &= \prod_{j=1}^d \mathbb{E}_0(1 + p_d(L_j - 1) - \mathbb{1}_{\overline{\mathcal{D}_j}}(p_d(L_j - 1) + 1)) \\ &= \prod_{j=1}^d \left(1 - p_d(\mathbb{E}_0(L_j \mathbb{1}_{\overline{\mathcal{D}_j}})) + (-1 + p_d)\mathbb{P}_0(\overline{\mathcal{D}_j})\right) \\ &= \exp\left(\sum_{j=1}^d \log\left(1 - p_d(\mathbb{E}_0(L_j \mathbb{1}_{\overline{\mathcal{D}_j}})) + (-1 + p_d)\mathbb{P}_0(\overline{\mathcal{D}_j})\right)\right), \end{aligned} \quad (7.50)$$

where $\mathcal{D}_j = \{\tilde{l}_j \leq a(r_\epsilon)\sqrt{(2+v)\log d}\}$ and $\overline{\mathcal{D}_j}$ denotes the complement of \mathcal{D}_j . Relation (7.46) entails the convergence to zero of the second term in the log term of the right-hand side of (7.50). Therefore, in order to obtain $\mathbb{E}_0(\hat{L}_{\Pi^d}) \rightarrow 1$, it is sufficient to prove that

$$dp_d(\mathbb{E}_0(L_j \mathbb{1}_{\overline{\mathcal{D}_j}})) = o(1). \quad (7.51)$$

Note that $\mathbb{E}_0(L_j \mathbb{1}_{\overline{\mathcal{D}_j}}) = \mathbb{P}_{\nu_j}(\overline{\mathcal{D}_j})$. Since $\frac{\sqrt{2+v}}{c(r_\epsilon)} - 1$ is positive ($c(r_\epsilon) < \sqrt{2}$) for any positive v , we can applied relation (7.48) of Lemma 7.3 to get

$$\begin{aligned} dp_d \mathbb{P}_{\nu_j}(\overline{\mathcal{D}_j}) &\leq dp_d \exp\left(-\frac{1}{2} \log(d)((\sqrt{2+v} - c(r_\epsilon))^2 + o(1))\right) \\ &= d^{1-b}(1 + \rho_d) d^{-\frac{1}{2}(\sqrt{2+v} - c(r_\epsilon))^2 + o(1)}, \end{aligned} \quad (7.52)$$

where the right-hand side of (7.52) goes to zero as soon as $c(r_\epsilon) < \sqrt{2+v} - \sqrt{2(1-b)}$. This yields relation (7.51).

Second, we need to study $\mathbb{E}_0(\hat{L}_{\Pi^d}^2)$:

$$\begin{aligned} \mathbb{E}_0(\hat{L}_{\Pi^d}^2) &= \prod_{j=1}^d \mathbb{E}_0((1 - p_d(1 - L_j))^2 \mathbb{1}_{\mathcal{D}_j}) \\ &= \exp\left(\sum_{j=1}^d \log(1 - 2p_d \mathbb{E}_0((1 - L_j) \mathbb{1}_{\mathcal{D}_j}) + \mathbb{E}_0(p_d^2(1 - L_j)^2 \mathbb{1}_{\mathcal{D}_j} - \mathbb{1}_{\overline{\mathcal{D}_j}}))\right). \end{aligned}$$

Since the relations $d\mathbb{P}_0(\overline{\mathcal{D}_j}) = o(1)$ and $dp_d \mathbb{E}_0((1 - L_j) \mathbb{1}_{\mathcal{D}_j}) = o(1)$ have been already proved, it is sufficient to show that $dp_d^2 \mathbb{E}_0((1 - L_j)^2 \mathbb{1}_{\mathcal{D}_j}) = o(1)$. To this end, observe that

$$d\mathbb{E}_0(p_d^2(1 - L_j)^2 \mathbb{1}_{\mathcal{D}_j}) \leq 2(dp_d^2 \mathbb{P}_0(\mathcal{D}_j) + dp_d^2 \mathbb{E}_0(L_j^2 \mathbb{1}_{\mathcal{D}_j})). \quad (7.53)$$

The first term on the right-hand side of (7.53) tends to zero as d goes to infinity since $dp_d^2 = dd^{-2b}(1 + \rho_d)^2$ for $b \in (3/4, 1)$.

To study of the second term on the right-hand side of (7.53), we take into account the following two points:

(i) since $\sup_k z_k^2 = o(1)$, we can apply Lemma 7.4 of Section 7.4 with $h = 2$, $X = x_{j,k}/\epsilon$, and $z = z_k$, and obtain

$$\exp(2\tilde{l}_j) = \exp(2a^2(r_\epsilon)). \quad (7.54)$$

(ii) since $\limsup(T/a^2(r_\epsilon)) < 2$ is satisfied as soon as $c(r_\epsilon) > \frac{\sqrt{2}}{2}$ with $T = a(r_\epsilon)\sqrt{(2+v)\log d}$, we can applied relation (7.49) of Lemma 7.3, which jointly with relation (7.54) leads to

$$\begin{aligned}
dp_d^2 \mathbb{E}_0(L_j^2 \mathbb{1}_{\mathcal{D}_j}) &= dd^{-2b}(1+\rho_d)^2 \mathbb{P}_{\mu_j}(\tilde{l}_j \leq a(r_\epsilon)\sqrt{\log d} \sqrt{(2+v)}) \exp(2\tilde{l}_j - a^2(r_\epsilon)) \\
&< dd^{-2b}(1+\rho_d)^2 \times \\
&\quad \exp\left(-\frac{a^2(r_\epsilon)(\sqrt{2+v}\sqrt{\log d} - 2a(r_\epsilon))^2}{2a^2(r_\epsilon)} + a^2(r_\epsilon) + o(a^2(r_\epsilon))\right) \\
&= dd^{-2b}(1+\rho_d)^2 \exp\left(-\frac{\log d}{2}((\sqrt{2+v} - 2c(r_\epsilon))^2 + c^2(r_\epsilon) + o(1))\right) \\
&= dd^{-2b}(1+\rho_d)^2 d^{-\frac{1}{2}(\sqrt{(2+v)} - 2c(r_\epsilon))^2 + c^2(r_\epsilon) + o(1)}. \tag{7.55}
\end{aligned}$$

The expression on the right-hand side of (7.55) goes to zero as soon as $c(r_\epsilon) < \sqrt{2+v} - \sqrt{2(1-b)}$. The last inequality is obtained by resolving the inequality $1 - 2b - \frac{1}{2}(\sqrt{2+v} - 2x)^2 + x^2 < 0$, where x is constrained to be larger than $\frac{\sqrt{2}}{2}$. This implies that a successful detection is impossible as soon as $c(r_\epsilon) < \varphi_2(b)$, where φ_2 is defined by (2.2).

7.4. Appendix

7.4.1. Proof of Lemma 7.2.

If there exists λ such that (7.19) is valid, then equation (7.20) is obtained in adapting Lemma 7.4.'s proof of [17]. Indeed, due to (7.19) and using the fact that $\sum_{j=1}^K \eta_j \geq K\eta_0$, we obtain for all $\eta_j \in [0, R]$, $j \in \{1, \dots, K\}$:

$$\begin{aligned}
\sum_{j=1}^K f_T(\eta_j) &\geq \sum_{j=1}^K \inf\{f_T(\eta_j) - \lambda\eta_j\} + \lambda K\eta_0 \\
&\geq K(f_T(\eta_0) - \lambda\eta_0) + \lambda K\eta_0 \\
&= Kf_T(\eta_0). \tag{7.56}
\end{aligned}$$

On the other hand,

$$\begin{aligned}
F_{K,T}(\eta_0) &= \inf_{\{(\eta_1, \dots, \eta_K) : \sum \eta_j \geq K\eta_0\}} \sum_{j=1}^K f_T(\eta_j) \\
&\leq Kf_T(\eta_0). \tag{7.57}
\end{aligned}$$

Relations (7.56) and (7.57) yield relation (7.20).

Now, let us prove that (7.18) implies (7.19). For this, set $g_T(\eta) = f_T(\eta) - \lambda\eta$ and denote by g'_T and $g_T^{(2)}$ the first and second derivatives of g_T , respectively. Note that $g'_T(\eta) = (T - \eta)f_T(\eta) - \lambda$, and we choose $\lambda = (T - \eta_0)f_T(\eta_0)$ to have $g'_T(\eta_0) = 0$.

The study of $g_T^{(2)}$ yields that $g_T^{(2)} > 0$ for $|T - \eta| > 1$ and $g_T^{(2)} < 0$ for $|\eta - T| < 1$. Since $0 < \eta_0 < T - 1$, this implies that $g'_T < 0$ on $[0, \eta_0]$, $g'_T(\eta_0) = 0$, $g'_T > 0$ on $[\eta_0, T - 1]$, g'_T is decreasing on $]T - 1, T + 1]$, and g'_T is increasing on $]T + 1, +\infty[$. Moreover, $g'_T(T - 1) > 0$ and $g'_T(T) = -\lambda < 0$, so that there exists $t \in]T - 1, T[$ such that $g'_T(t) = 0$. This yields that η_0 is a local minimum of g_T . In order to prove that η_0 is a global minimum of g_T , it is sufficient to show that $g_T(R) - g_T(\eta_0) > 0$. Let us set $R = T + x$, with a positive real x . We already know that $x < T - \eta_0$ since $g_T(T + (T - \eta_0)) = f_T(\eta_0) - \lambda(T + T - \eta_0) = f_T(\eta_0) - \lambda\eta_0 - 2\lambda(T - \eta_0) < g_T(\eta_0)$, where the last inequality is valid because of the choice of λ and $T - \eta_0$. For $x < (T - \eta_0)$, we obtain

$$\begin{aligned}
g_T(R) - g_T(\eta_0) &= \exp\left(-\frac{x^2}{2}\right) - (T - \eta_0)f_T(\eta_0)(T + x) - f_T(\eta_0) + (T - \eta_0)f_T(\eta_0)\eta_0 \\
&> \exp\left(-\frac{x^2}{2}\right) - f_T(\eta_0)(2(T - \eta_0)^2 + 1) > 0, \tag{7.58}
\end{aligned}$$

where inequality (7.58) is valid as soon as

$$\exp(-\frac{x^2}{2}) > \exp(-\frac{(T-\eta_0)^2}{2})(2(T-\eta_0)^2+1) \Leftrightarrow x < ((T-\eta_0)^2 - 2\log(2(T-\eta_0)^2+1))^{1/2}.$$

Since (7.18) implies (7.19), this completes the proof of Lemma 7.2.

7.4.2. Proof of Lemma 7.3.

The proof of Lemma 7.3 requires an additional result stated as Lemma 7.4 below. For any $j \in \{1, \dots, d\}$, recall that l_j and \tilde{l}_j are given by (7.42) and (7.41), respectively. For any $j \in \{1, \dots, d\}$ and $k \in \mathbf{Z}$, set $\tilde{l}_{j,k} = \frac{z_k^4}{4} - \frac{z_k^2}{2} + \log(\cosh(z_k x_{j,k}/\epsilon))$ and $l_{j,k} = -\frac{z_k^2}{2} + \log(\cosh(z_k x_{j,k}/\epsilon))$. Denote by Λ_j , $\tilde{\Lambda}_j$, and $\tilde{\Lambda}_{j,k}$ the moment-generating functions of l_j , \tilde{l}_j , and $\tilde{l}_{j,k}$ under \mathbb{P}_0 , respectively. From equations (7.35) and (7.41), it turns out that for any h ,

$$\tilde{\Lambda}_j(h) = \prod_{k \in \mathbf{Z}} \tilde{\Lambda}_{j,k}(h), \quad (7.59)$$

$$\tilde{\Lambda}_j(h) = \Lambda_j(h) \exp(h \frac{a^2(r_\epsilon)}{2}). \quad (7.60)$$

Next, define the function $\tilde{g} : (z, y) \rightarrow \frac{z^4}{4} - \frac{z^2}{2} + \log(\cosh(zy))$, and observe that the following relations hold:

$$\begin{aligned} \tilde{l}_{j,k} &= \tilde{g}(z_k, x_{j,k}/\epsilon), \\ \tilde{\Lambda}_{j,k}(h) &= \mathbb{E}_0(\exp(h\tilde{g}(z_k, x_{j,k}/\epsilon))). \end{aligned} \quad (7.61)$$

Lemma 7.4. *Let X be a real standard Gaussian random variable. For any $z = o(1)$ and any $h = O(1)$,*

$$\log(\mathbb{E}(\exp(h\tilde{g}(z, X)))) = h^2 \frac{z^4}{4} + o(z^4).$$

Proof of Lemma 7.4.

For some $\delta > 0$, consider the event $\mathcal{E} = \{|zX| < \delta\}$ and denote by $\bar{\mathcal{E}}$ its complement in \mathbf{R} . We shall study the expectations $G_1(h, \delta) = \mathbb{E}(\exp(h \log(\cosh(zX))) \mathbb{1}_{\mathcal{E}})$ and $G_2(h, \delta) = \mathbb{E}(\exp(h \log(\cosh(zX))) \mathbb{1}_{\bar{\mathcal{E}}})$ separately. At this point, we choose δ small enough ($\delta = o(1)$) to satisfy $z\delta^{-1} = o(1)$.

First, let us study the term $G_2(h, \delta)$. With the use of the inequality $\cosh(x) \leq \exp(|x|)$, $\forall x \in \mathbf{R}$, and the fact that $h = O(1)$, the routine calculations of exponential moments of a real Gaussian random variable lead to

$$\begin{aligned} G_2(h, \delta) &\leq \mathbb{E}(\exp(h|zX|) \mathbb{1}_{(|zX| \geq \delta)}) \\ &= 2\mathbb{E}(\exp(hzX) \mathbb{1}_{(X \geq \delta/z)}) \\ &= \frac{2}{\sqrt{2\pi}} \int_{\mathbf{R}^+} \exp(-\frac{1}{2}(x - hz)^2) \mathbb{1}_{(x \geq \frac{\delta}{z})} dx \exp(\frac{1}{2}h^2 z^2) \\ &\leq 2 \exp(h^2 \frac{z^2}{2}) \exp(-\frac{1}{2}(\frac{\delta}{z} - hz)^2) \\ &\leq 2 \exp\left(-\frac{1}{2} \frac{\delta^2}{z^2} + o(1)\right), \end{aligned} \quad (7.62)$$

where, with our choice of δ , the right-hand side of (7.62) is small.

Now, we move on to the term $G_1(h, \delta)$. If δ is small enough, then $|zX|$ is also small and the following relation holds:

$$\log(\cosh(zX)) = \frac{z^2}{2} X^2 - \frac{z^4}{12} X^4 + o(z^4 X^4). \quad (7.63)$$

Then the routine calculations of exponential moments as above lead to the following:

$$\begin{aligned}
G_1(h, \delta) &= \mathbb{E} \left(\exp \left(h \left(\frac{z^2}{2} X^2 - \frac{z^4}{12} X^4 (1 + o(1)) \right) \right) \mathbb{I}_{\mathcal{E}} \right) \\
&= \mathbb{E} \left(\exp \left(h \left(\frac{z^2}{2} X^2 \right) \right) \left(1 - h \frac{z^4}{12} X^4 (1 + o(1)) \right) \mathbb{I}_{\mathcal{E}} \right) \\
&= \exp \left(-\frac{1}{2} \log(1 - h z^2) \right) \exp \left(-\frac{h}{4} z^4 (1 + o(1)) \right) \\
&= \exp \left(\frac{h}{2} z^2 + \frac{h^2}{4} z^4 (1 + o(1)) \right) \exp \left(-\frac{h}{4} z^4 (1 + o(1)) \right) \\
&= \exp \left(\frac{h}{2} z^2 + \frac{h^2}{4} z^4 - \frac{h}{4} z^4 + o(z^4) \right). \tag{7.64}
\end{aligned}$$

Taking $h = O(1)$, $z = o(1)$, $\delta = o(1)$ and $z\delta^{-1} = o(1)$ in relations (7.62) and (7.64) entails that $G_1(h, \delta) = O(1)$, $G_2(h, \delta) = O(\exp(-\delta^2/(2z^2))) = o(1)$, and therefore $G_2(h, \delta)(G_1(h, \delta))^{-1} = o(1)$.

Next, due to (7.62), (7.64) and using the fact that $h = O(1)$, $z = o(1)$, for small δ such that $z_0\delta^{-1} = o(1)$ and $\delta = o(1)$, we obtain

$$\begin{aligned}
\log(\mathbb{E}(\exp(h\tilde{g}(z, X)))) &= \log(G_1(h, \delta) + G_2(h, \delta)) - \frac{h}{2} \left(z^2 - \frac{z^4}{2} \right) \\
&= \left(\log G_1(h, \delta) - \frac{h}{2} \left(z^2 - \frac{z^4}{2} \right) \right) + \log \left(1 + \frac{G_2(h, \delta)}{G_1(h, \delta)} \right) \\
&= h^2 \frac{z^4}{4} + o(z^4) + \frac{G_2(h, \delta)}{G_1(h, \delta)} (1 + o(1)) \\
&= \left(h^2 \frac{z^4}{4} + o(z^4) \right) \left(1 + \frac{G_2(h, \delta)(1 + o(1))}{G_1(h, \delta)(h^2 \frac{z^4}{4} + o(z^4))} \right) \\
&= h^2 \frac{z^4}{4} + o(z^4), \tag{7.65}
\end{aligned}$$

where relation (7.65) holds provided that

$$\frac{G_2(h, \delta)}{G_1(h, \delta)(h^2 \frac{z^4}{4} + o(z^4))} = o(1). \tag{7.66}$$

It is then sufficient to prove (7.66) since (7.65) is the expected result of Lemma 7.4. Recall that $h = O(1)$ and $z = o(1)$ entail that $G_1(h, \delta) = O(1)$ and $G_2(h, \delta) = O(\exp(-\delta^2/(2z^2)))$. Then, it is sufficient to establish that $\exp(-\frac{1}{2} \frac{\delta^2}{z^2}) z^{-4} = o(1)$. The latter holds if we choose δ such that $\delta^{-1} = o((z \sqrt{\log(z^{-1})})^{-1})$.

Proof of Lemma 7.3.

Remark 4.2 and relations (4.10) imply that $\sup_k z_k^2 \leq z_0^2 = o(1)$ as soon as $\log(d) = o(\epsilon^{-2/(2\tau+1)})$.

Due to (7.61), for any h such that $h = O(1)$, Lemma 7.4 can be applied to the moment-generating function $\Lambda_{j,k}(h)$.

Here and later, we consider any $j \in \{1, \dots, d\}$ and any $k \in \mathbf{Z}$. Due to relations (7.59), (7.61), (7.21), (7.41), by applying Lemma 7.4 and using the exponential Chebyshev's inequality, we obtain for any positive h such that $h = O(1)$,

$$\begin{aligned}
\mathbb{P}_0(\tilde{l}_j > T) &\leq \tilde{\Lambda}_j(h) \exp(-hT) \\
&\leq \exp \left(\frac{h^2}{2} a^2(r_\epsilon) - hT + o(a^2(r_\epsilon)) \right). \tag{7.67}
\end{aligned}$$

The minimum on the right-hand side of (7.67) is attained for $h = \frac{T}{a^2(r_\epsilon)}$ which is positive and of order 1; this allows us to prove relation (7.47).

Due to relations (7.61), (7.59), (7.21), (7.41), (7.43), by applying again Lemma 7.4 and using the exponential Chebyshev's inequality, we obtain for any positive h such that $h = O(1)$,

$$\begin{aligned} \mathbb{P}_{\nu_j}(\tilde{l}_j > T) &\leq \mathbb{E}_{\nu_j}(\exp(\tilde{l}_j h)) \exp(-hT) \\ &= \tilde{\Lambda}_j(h+1) \exp\left(-\frac{a^2(r_\epsilon)}{2} - hT\right) \\ &= \exp\left(\frac{(h+1)^2}{2} a^2(r_\epsilon) - \frac{a^2(r_\epsilon)}{2} - hT + o(a^2(r_\epsilon))\right), \end{aligned} \quad (7.68)$$

where the minimum in the right-hand side of (7.68) is attained for $h = \frac{T}{a^2(r_\epsilon)} - 1$ which is positive and of order 1; this yields relation (7.48).

Recall that under the assumption of Lemma 7.3, the quantity $2a^2(r_\epsilon) - T$ is positive. Therefore, from (7.61), (7.59), (7.21), (7.44), (7.41), and (7.60), applying Lemma 7.4 and using the exponential Chebyshev's inequality, we get for any positive h such that $h = O(1)$,

$$\begin{aligned} \mathbb{P}_{\mu_j}(\tilde{l}_j \leq T) &= \mathbb{P}_{\mu_j}(-\tilde{l}_j \geq -T) \\ &= \mathbb{E}_{\mu_j}(\exp(-\tilde{l}_j h)) \exp(hT) \\ &= \mathbb{E}_0(\exp(-\tilde{l}_j h) \exp(2\tilde{l}_j)) \exp(-a^2(r_\epsilon)) (\Lambda_j(2))^{-1} \exp(hT) \\ &= \tilde{\Lambda}_j(2-h) (\Lambda_j(2))^{-1} \exp(-a^2(r_\epsilon) + Th) \\ &= \tilde{\Lambda}_j(2-h) (\tilde{\Lambda}_j(2))^{-1} \exp(a^2(r_\epsilon)) \exp(-a^2(r_\epsilon) + Th) \\ &= \exp\left(\frac{1}{2}(2-h)^2 a^2(r_\epsilon) - 2a^2(r_\epsilon) + Th + o(a^2(r_\epsilon))\right), \end{aligned} \quad (7.69)$$

where the minimum in the right-hand side of (7.69) is achieved for $h = -\frac{T}{a^2(r_\epsilon)} + 2$ which is positive and of order $O(1)$; this yields relation (7.49). The proof of Lemma 7.3 is completed.

Acknowledgements: We thank one of our colleagues for his contribution to improve the English language of the paper.

References

- [1] BICKEL, P.J., RITOV, Y. and TSYBAKOV, A.B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732.
- [2] CAI, T., JIN, J. and LOW, M. (2007). Estimation and confidence sets for sparse normal mixtures. *Ann. Statist.* **35** 2421–2449.
- [3] DONOHO, D.L. (2006). Compressed Sensing. *IEEE Transactions on Information Theory* **52** 1289–1306.
- [4] DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994.
- [5] HUANG, J., HOROWITZ, J.L. and WEI, F. (2010). Variable selection in nonparametric additive models. *Ann. Statist.* **38** 2282–2313.
- [6] IBRAGIMOV, I.A. and KHASHINSKII, R.Z. (1997). Some estimation problems on Infinite dimensional Gaussian white noise. In: *Festschrift for Lucien Le Cam. Research papers in Probability and Statistics*, 275–296. Springer-Verlag, New York.
- [7] INGSTER, YU.I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods of Statist.* **6** 47–69.
- [8] INGSTER, YU.I. (2001). Adaptive detection of a signal of growing dimension. I. *Math. Methods Statist.* **10** 395–421.
- [9] INGSTER, YU.I. (2002). Adaptive detection of a signal of growing dimension. II. *Math. Methods Statist.* **11** 37–68.

- [10] INGSTER, YU.I. and LEPSKI, O. (2003). On multichannel signal detection. *Math. Methods Statist.* **12** 247–275.
- [11] INGSTER, YU.I., POUET, CH. and TSYBAKOV, A.B. (2009). Classification of sparse high-dimensional vectors. *Phi. Trans. R. Soc. A.* **367** 4427–4448.
- [12] INGSTER, YU.I. and SUSLINA, I.A. (2002). On a detection of a signal of known shape in multichannel system. *Zapiski Nauchn. Sem. POMI* **294** 88–112 (Transl. *J. Math. Sci.* (2005) **127** 1723–1736).
- [13] INGSTER, YU.I. and SUSLINA, I.A. (2003). *Nonparametric goodness-of-fit testing under gaussian models*. Lectures Notes in Statistics. vol. 169., Springer-Verlag, New York.
- [14] INGSTER, YU.I. and SUSLINA, I.A. (2005). On estimation and detection of smooth function of many variables. *Math. Methods Statist.* **14** 299–331.
- [15] INGSTER, YU.I. and SUSLINA, I.A. (2007). Estimation and detection of high-variable functions from Sloan-Woźniakowski space. *Math. Methods Statist.* **16** 318–353.
- [16] INGSTER, YU.I. and SUSLINA, I.A. (2007). On estimation and detection of a function from tensor product spaces. (In Russian). *Zapiski Nauchn. Sem. POMI* **351** 180–218. (Translation in: *J. Math. Sci.* (2008), **152** 897–920).
- [17] INGSTER, YU.I., TSYBAKOV, A.B. and VERZELEN, N. (2010). Detection boundary in sparse regression. *Electronic J. of Statistics* **4** 1476–1526.
- [18] LIN, Y. (2000). Tensor product space ANOVA model. *Ann. Statist.* **28** 734–755.
- [19] RASKUTTI, G., WAINWRIGHT, M.J. and YU, B. (2011). Minimax-optimal rates for sparse additive models over kernel classes via convex programming.
<http://arxiv.org/abs/1008.3654>
- [20] STONE, CH. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13** 689–705.
- [21] TIBSHIRANI, R. (1996). Regression shrinkage and selection via Lasso. *J. Roy. Statist. Soc. Ser. B.* **58** 267–288.